PL-TR-93-2202

PHILLIPS LABORATORY
GEOPHYSICS SCHOLAR PROGRAM

Editors:

Warren D. Peele
Earl L. Steele

Southeastern Center for Electrical Engineering Education, Inc.
(SCEEE)
1101 Massachusetts Avenue
St. Cloud, FL 34769

DTIC
ELECTE
DEC 06 1993
S D
E

30 September 1993

Final Report
7 October 1986 - 30 September 1993

PHILLIPS LABORATORY
Directorate of Geophysics
AIR FORCE MATERIEL COMMAND
HANSCOM AFB, MA 01731-3010

93 12 3 045

93-29602

Best
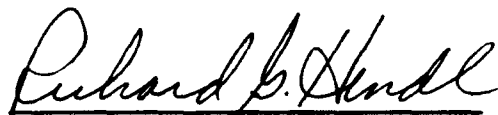# Available
Copy

"This technical report has been reviewed and is approved for publication"

Adolph S. Jursa

ADOLPH S. JURSA
Contract Manager

Richard G. Hendl

RICHARD G. HENDL
Chief Scientist, Geophysics

This report has been reviewed by the ESC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center (DTIC). All others should apply to the National Technical Information Service (NTIS).

If your address has changed, or if you wish to be removed from the mailing list, of if the addressee is no longer employed by your organization, please notify PL/TSI, 29 Randolph Road, Hanscom AFB, MA 01731-3010. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget. Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 30 September 1993	3. REPORT TYPE AND DATES COVERED Final (7 Oct 1986–30 Sep 1993)

4. TITLE AND SUBTITLE
Phillips Laboratory Geophysics Scholar Program

5. FUNDING NUMBERS
PE 62101F
PR 9993
TA GS
WU PR

Contract F19628-86-C-0224

6. AUTHOR(S)
Editors:
Warren D. Peele
Earl L. Steele

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Southeastern Center for Electrical Engineering Education
(SCEEE)
1101 Massachusetts Avenue
St Cloud, FL 34769

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Phillips Laboratory
29 Randolph Road
Hanscom AFB, MA 01731-3010

Contract Manager: Adolph Jursa/CAG

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

PL-TR-93-2202

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Management and Technical Report describing Phillips Laboratory Geophysics Scholar Program at the Geophysics Directorate.

Forty-two (42) Phillips Laboratory Scholars have been active in the exploratory and advanced development for one (1) year research periods in those areas of geophysics which meet known and anticipated military requirements. The Final Summary of that effort appears in this report along with the Technical Reports of thirteen (13) scholars.

14. SUBJECT TERMS
Geophysics, Atmospheric Science, Meteorology, Space Physics,
Molecular Physics, Mathematics, Chemistry, Computer Science

15. NUMBER OF PAGES
438

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

INTRODUCTION AND SUMMARY

The Phillips Laboratory Scholar Program was initiated in June of 1982 as the "Air Force Geophysics Scholar Program". It was implemented as an addition to the Air Force Summer Faculty Program contract and was addressed to emerging post-doctoral researchers who had recently completed the Ph.D. The program was transitioned to a separate contract for the period 1983-1986.

The program discussed in this report ran from 7 October 1986 - 30 September 1993. The Scholars have been active in exploratory and advanced development in those areas of geophysics which meet known and anticipated military requirements. Six annual technical reports have been published describing the research effort accomplished. These reports are on file in the office of the chief scientist of the Phillips Laboratory Geophysics Directorate.

Scholar research activity during the program was impressive. There were 42 scholars participating and they published 71 papers in professional journals, made approximately 90 oral presentations at professional meetings, and each produced a final report. The reports for the final year are bound in this volume.

The program description which follows is taken from the advertising brochure used during the last two years of the contract.

PROGRAM DESCRIPTION

The Phillips Laboratory Geophysics Directorate has initiated the Geophysics Scholar Program to broaden the direct participation of qualified researchers in research programs. The program provides research opportunities for selected Engineers and Scientists holding a doctoral degree to work at the Phillips Laboratory Geophysics Directorate for a 12 month research period. The appointment may be extended for a second term.

To be eligible, all candidates must have a Ph.D. or equivalent experience in an appropriate technical field. The Scholars will be selected primarily from such basic and applied science fields as physics particularly geophysics and atmospheric physics, meteorology, and space physics, and also from mathematics, molecular physics, chemistry, computer science, and engineering. Applicants must be U.S. citizens.

The Air Force Geophysics Scholar in this program will have the following specific obligations:

1) To participate in advanced research programs at the Phillips Laboratory Geophysics Directorate;

2) To prepare a report at the end of the appointment describing the research accomplishments. This report will be subject to approval by the Air Force Geophysics Directorate.

Application Information

Qualified technical people who are interested in an appointment under this program should file a formal application and supporting materials with the program director at the address noted on the front page. Formal application forms are included with this booklet. If additional forms are

needed, they may be obtained from the SCEEE programs office. Appointments may be made at any time, thus prompt application is encouraged.

SCEEE supports equal opportunity/affirmative action. All qualified applicants will receive consideration without regard to race, color, religion, sex, or national origin.

Phillips Laboratory Geophysics Scholar Program Objectives

1) To provide a productive means for scientists and engineers holding Ph.D. degrees to participate in research at the Air Force Geophysics Directorate;

2) To stimulate continuing professional association among the Scholars and their professional peers in the Air Force;

3) To further the research objectives of the United States Air Force; and

4) To enhance the research productivity and capabilities of scientists and engineers especially as these relate to Air Force technical interests.

Prerequisites for Appointments

To be qualified for consideration as a Geophysics Scholar the applicant must:

1) Be a U.S. citizen;

2) Be the holder of a Ph.D. degree, or equivalent, in an appropriate technical speciality; and

3) Be willing to pursue research work of limited time duration at the Air Force Geophysics Directorate.

Although it is anticipated that the research itself may be unclassified, the Scholar must hold or be eligible for a Department of Defense SECRET clearance in order to insure access to work areas.

Research Period

The period of this appointment is for 12 months at the Phillips Laboratory Geophysics Directorate research site, Hanscom AFB, Massachusetts. Appointments may be made at any time.

Financial Terms

Participants are paid a negotiated salary, plus a 10% salary increment payable at the completion of the program, 26 days vacation, and 13 days of sick leave per year. Unused vacation and sick leave will be reimbursed at the end of the year.

Travel

Travel expenses will be reimbursed for one trip from the Scholar's normal location to the Air Force facility at the start of the appointment; and one return trip from the Laboratory to the Scholar's normal home base at the end of the appointment period. One pre-appointment visit may also be authorized. This travel will be reimbursed in accordance with SCEEE travel policy. Costs of relocation of professional equipment and household goods will be reimbursed if appropriate.

PARTICIPATING SCHOLARS

The following list includes all participating scholars with geographical and educational background data.

<u>Geophysics Scholars</u>

Harro Ackermann Albuquerque, NM	University of New Mexico Ph.D., Physics (Optics), 1985
Jay Albert Ithaca, NY	Princeton University Ph.D., Astrophysical Sciences (Plasma Physics), 1986
Stephen Anderson Somerville, MA	Princeton University Ph.D., Plasma Physics (Plasma Astrophysics), 1992
Peter Armstrong Charlottesville, VA	St. Lawrence University Ph.D., Physics (Atomic Physics), 1991
Rodger Biasca Cambridge, MA	Massachusetts Institute of Technology Ph.D., Aeronautics/Astronautics (Space Physics), 1991
Steve Cardimona Austin, TX	University of Texas at Austin Ph.D., Geophysics (Seismology), 1992
Robert Cicerone Dedham, MA	Massachusetts Institute of Technology Ph.D., Geophysics (Geophysics), 1991
Anthony Dentamaro Medford, MA	Tufts University Ph.D., Physics (Perturbative Quantum Chromodynamics), 1985
James Dodd Cambridge, MA	Stanford University Ph.D., Chemistry (Physical Chemistry), 1986
Hoang Dothe Boston, MA	The Johns Hopkins University Ph.D., Chemistry (Theoretical Chemistry), 1987
Richard Eastes Maynard, MA	The Johns Hopkins University Ph.D., Physics (Physics), 1985
Gary Erickson Greenbelt, MD	Rice University Ph.D., Space Physics (Magnetospheric Physics), 1985
Robert Farley State College, PA	Univ. of Colorado Boulder Ph.D., Chemical Physics (Physical Chemistry), 1989
Dorothy Flanagan Acton, MA	California Institute of Technology Ph.D., Chemistry (Chemical Physics), 1985

John Foley
Lowell, MA

Massachusetts Institute of Technology
Ph.D., Geophysics (Seismic Tomography & Inversion), 1990

James Gardner, II
Hudson, NH

Boston College
Ph.D., Chemistry (Atmospheric Chemistry), 1988

Gregory Ginet
Ithaca, NY

Cornell University
Ph.D., Applied Physics (Theoretical Plasma Physics), 1987

Richard Gross
Santa Fe, NM

University of Colorado at Boulder
Ph.D., Geophysics (Solid Earth Geophysics), 1982

Mark Handel
Jamaica Plain, MA

Massachusetts Institute of Technology
Sc.D., Atmospheric Physics, (Meteorology/Atmospheric Physics), 1991

Robert Hawkins
Washington, DC

Ohio State University
Ph.D., Physics (Spectroscopy), 1981

Stan Heckman
Cambridge, MA

Massachusetts Institute of Technology
Ph.D., Atmospheric Science (Atmospheric Electricity), 1991

Michael Hoke
New Carlisle, OH

Ohio State University
Ph.D., Physics (Physics), 1982

Samuel Howard
Leakesville, MS

Florida State University
Ph.D., Physics (Optics - Imaging Processing), 1982

Paul Keyser
Boulder, CO

University of Colorado - Boulder
Ph.D., Physics (Experimental Gravitational Physics), 1986

Juergen Krause-Polstorff
Monterey, CA

Rice University
Ph.D., Space Physics and Astronomy (Theoretical Astrophysics), 1983

Thomas Kuchar
Watertown, MA

Boston University
Ph.D., Astronomy and Physics (Interstellar Medium), 1992

Usama Makhlouf
Bedford, MA

University of Cincinnati
Ph.D., Space Physics (Gravity Waves), 1989

Robert McCaffrey
Braintree, MA

University of California, Santa Cruz
Ph.D., Geophysics (Seismology), 1981

Martin McHugh
Boulder, CO

University of Colorado
Ph.D., Physics (Experimental Gravitational Physics), 1991

Robert Morris
Brighton, MA

Boston University
Ph.D., Physical Chemistry (Gas Phase Chemical Kinectics and Spectroscopy), 1987

4

Bao Nguyen
St. Louis, MA

St. Louis University
Ph.D., Geophysics (Seismology), 1988

Mark Popecki
Newmarket, NH

Univ. of New Hampshire
Ph.D., Physics (Magnetospheric Physics), 1991

Timothy Schumaker
Brighton, MA

Boston College
Ph.D., Physics, (Space Physics), 1986

Gregory Sloan
Laramie, WY

Univ. of Wyoming
Ph.D., Physics (Astrophysics), 1992

Amy Stevens-Miller
Norman, OK

California Institute of Technology
Ph.D., Chemistry (Physical Chemistry), 1981

Joseph Thomas
Pittsburgh, PA

University of Pittsburgh
Ph.D., Chemistry (Gas Phase Chemical Kinetics), 1986

Jane VanDoren
Cambridge, MA

University of Colorado
Ph.D., Chemical Physics, (Gas Phase Reaction Kinetics,
    Thermodynamics, and Mechanisms), 1987

Donald Vasco
Berkeley, CA

University of California at Berkeley
Ph.D., Geophysics (Extremal Inversion of Earth
    Displacement), 1986

Terri Vossler
Austin, TX

Washington University
Ph.D., Chemistry (Atmospheric Sciences), 1985

Lyn Watson
Medford, MA

Boston College
Ph.D., Chemistry (Physical Chemistry), 1989

Raymond Willemann
Arcadia, CA

Cornell University
Ph.D., Geophysics (Lithospheric Flexure), 1986

Lorraine Wolf
Fairbanks, AK

University of Alaska, Fairbanks
Ph.D., Geophysics (Seismology), 1989

SCHOLAR ACCOMPLISHMENTS

<u>**Publications**</u>

Abers, G., R. McCaffrey, 1988, Active deformation in the New Guinea Fold-and-Thrust Belt: Seismological Evidence for Strike-slip Faulting and Basement-involved Thrusting, *J. Geophys. Res.*, in press.

R. Biasca, D. Hastings, and D. Cooke, Simulation of the Critical Ionization Velocity: Effect of Using Physically Correct Mass Ratios, *J. Geophys Res.*, 97, 6451, 1992.

S. Cardimona, Integral Equations and Scatterer Imaging, submitted to *Inverse Problems.*

S. Cardimona, Waveform Inversion and Digital Filter Theory, *Geophysics*, 56, 534-536, 1991.

J. A. Dodd, W. A. M. Blumberg, S. J. Lipson, J. R. Lowell, P. S. Armstrong, D. R. Smith, R. M. Nadile, N. B. Wheeler, and E. R. Huppi, CIRRIS 1A and EXCEDE III Measurement of Column Densities of Rotationally Excited Nitric Oxide, *Geophys. Res. Lett.*, 20, 305, 1993.

J. A. Dodd, S. J. Lipson, and W. A. M. Blumberg, CIRRIS 1A and EXCEDE III Measurement of Column Densities of Rotationally Excited Nitric Oxide, *J. Chem. Phys.*, 95, 5752, 1991.

J. A. Dodd, S. J. Lipson, W. A. M. Blumberg, Formation and Vibrational Relaxation of $OH(X^2\Pi_i,v)$ by O_2 and Co_2, *J. Chem. Phys.*, 95, 5752-5762, 1991.

R. A. Dressler, J. A. Gardner, C. R. Lishawa, R. H. Salter, and E. Murad, Suprathermal Ion-Neutral Reactions: NO^+ Formation Cross Sections in N^+ -H_2O Collisions and OH Chemiluminescence in N_2^+ (Ar^+) -H_2O Collisions, *J. Chem. Phys.*, 93, 9189 1990.

R. A. Dressler, J. A. Gardner, R. H. Salter, F. J. Wodarczyk, and E. Murad, Measurements of Cross Sections and Product Ion Energies for Ion-Neutral Reactions at Suprathermal Energies, *J. Chem. Phys.*, in press 1989.

E. E. Ferguson, J.M. Van Doren, A. A. Viggiano, R. A. Morris, J. F. Paulson, Internal and translational energy effects on the charge-transfer reaction of CO_2 + with O_2, submitted to *Int. J. Mass Spectrom. Ion. Proc.*

J. A. Gardner, R. A. Dressler, R. H. Salter, and E. Murad, Suprathermal Ion-Neutral Reactions: NO^+ Formation Cross Sections in N^+ -H_2O Collisions and OH Chemiluminescence in N_2^+ (Ar^+) -H_2O Collisions, *J. Chem Phys.*, 93, 7780, 1990.

J. A. Gardner, R. A. Dressler, R. H. Salter, and E. Murad, Suprathermal Ion-Neutral Reactions: NO^+ Formation Cross Sections in N^+ -H_2O Collisions and OH Chemiluminescence in N_2^+ (Ar^+) -H_2O Collisions, *Chem Phys. Lett.*, submitted for publication.

M. D. Handel, J. S. Risbey, An Annotated Bibliography on the Greenhouse Effect and Climate Change, *Climatic Change*, 21, p. 97-255, 1992.

M. D. Handel, J. S. Risbey, P. H. Stone, Do We Know What Difference a Delay Makes?, *EOS, Transactions of the American Geophysical Union*, 72, (53), p. 596-597, 1991.

M. D. Handel, On the Asymmetry of Tropical Cyclone Outflows, *EOS, Transactions of the American Geophysical Union*, 73 (14, Spring Meeting Supplement), p. 73, 1992.

M. D. Handel, J. S. Risbey, Reflections on More than a Century of Climate Change Research, *Climatic Change*, 21, p. 91-96, 1992.

M. D. Handel, J. S. Risbey, P. H. Stone, Should We Delay Responses to the Greenhouse Issue?, *EOS, Transactions of the American Geophysical Union*, 72, (53), p. 593, 1991.

M. D. Handel, Subcritical Tropical Cyclone Intensification, in the preprint volume of the 8th Conference on Atmospheric and Oceanic Waves and Stability, *American Meteorological Society*, p. 315-319, 1991.

M. D. Handel, Tropical Cyclone Intensification. Part III: Ekman Pumping and the Formation of an Eye, submitted to the *Journal of the Atmopheric Sciences*.

M. D. Handel, Tropical Cyclone Intensification. Part I: Impediment to Linear Growth and Tropical Cyclone Intensification. Part II: Growth from Finite Amplitude Disturbances, submitted to the *Journal of the Atmospheric Sciences*.

Heinemann, M. and G. M. Erickson, Instability of sunward convection, submitted to *JGR*, 1989.

S. J. Howard, Fast Algorithm for Discrete Fourier Spectrum Extrapolation, *J. Opt. Soc. Am. A* 3 (13), p44 1986.

S. J. Howard, Fast Algorithm for Implementing the Minimum-Negativity Constraint for Fourier Spectrum Extrapolation, *Appl. Opt.*, 25, 1670 1986.

S. J. Howard, Fast Algorithm for Implementing the Minimum-Negativity Constraint for Fourier Spectrum Extrapolation II, *Appl. Opt.* 27, 3190 1988.

S. J. Howard, Minimum-Negativity Constraint Applied to Large Data Sets, *J. Opt. Soc. Am. A* 4(13), p22 1987.

D.H. Katayama and A. V. Dentamaro, Electronic Energy Transfer in Diatomic Molecules, *J. Chem. Phys.* 85, 2595 1986.

T. A. Kuchar, T. M. Bania, A High Resolution HI Survey of the Rosette Nebula, *Astrophysical Journal*.

R. Landau, G. Grasdalen, G. C. Sloan, Three-beam chopping: An Efficient Infrared Observing Technique, *Astron. Astrophys.* 259, 696-700, 1992.

P. D. LeVan, G.C. Sloan, I. R. Little-Marenin, Study of Infrared Emissions from H II Regions, *BAAS*, 25, 877, 1993.

P. D. LeVan, G. C. Sloan, I. R. Little-Marenin, G. L. Grasdalen, Study of Infrared Emissions from H II Regions, *ApJ*, 392, 702, 1992.

R. McCaffrey, Active tectonics of the eastern Sunda and Banda Arcs, *J. Geophys. Res.*, submitted, 1987a.

R. McCaffrey and J. Nabelek, Earthquakes, gravity and the origin of the Bali Basin: an example of a nascent continental fold-and-thrust belt, *J. Geophys. Res.*, 92, 441-460, 1987.

R. McCaffrey and J. Nabelek, Seismological evidence for shallow thrusting north of the Timor trough, *Geophys J. Roy. Astr. Soc.*, 85, 365-381, 1986.

J. McCarthy, Z. Hajnal, J. Howie, R. A. Johnson, G. R. Keller, L. W. Wolf, PACE lithospheric investigations of the Colorado Plateau, *EOS Trans.*, AGU, 71, 43, 1990.

M. P. McHugh, P. T. Keyser, J. E. Faller, New and Exoctic Phenomena '90 proceedings of the Xth Moriond Workshop, Les Arcs, France, 20-27 January 1990 *(Editions Fronti'eres, Gif Sur Yvette, 1990)*, Edited by O. Fackler and J. Tran Thanh Van, p. 233-6.

R. A. Morris, J. M. Van Doren, A. A. Viggiano and J. F. Paulson, Chemistry of C2- and HC2 - with Nitrogen, Oxygen, and Nitrogen Oxides, *J. Chem. Phys.*, Submitted, 1992.

R. A. Morris, A. A. Viggiano, F. Dale and J. F. Paulson, Chemistry of C2- and HC2- with Nitrogen, Oxygen, and Nitrogen Oxides, *J. Chem. Phys.*, 88, 1988, 4772.

R. A. Morris, A. A. Viggiano and J. F. Paulson, Internal and translational energy effects on the charge-transfer reaction of CO_2 + with O_2, *J. Chem. Phys.* 92, 2342 1990.

R. A. Morris, A. A. Viggiano and J. F. Paulson, Internal and translational energy effects on the charge-transfer reaction of CO_2 + with O_2, *J. Chem. Phys.*, 89, 4848 1990. *J. Chem. Phys.* 92, 3448 1990.

R. A. Morris, A. A. Viggiano and J. F. Paulson, Internal and translational energy effects on the charge-transfer reaction of CO_2 + with O_2, *J. Chem. Phys.* 94, 1884 1990.

T. L. Schumaker, D.A. Hardy, S. Moran, A. Huber, J. Pantazis, J. McGarity, Precipitating Electron and Ion Detectors (SSJ/4) for the Block 5D/Flight 8 DMSP Satellite, AFGL-TR-88-0030, 1988, ADA203990.

G. C. Sloan, P. D. LeVan, I. R. Little-Marenin, G. L. Grasdalen, Micron Spectroscopy of Carbon Stars Associated with Silicate Dust, *The Astrophysical Journal*, 392: 702-705, 1992.

G. C. Sloan, P. D. LeVan, P. C. Tandy, Report of Operations of the Air Force Geophysics Laboratory Infrared Array Spectrometer, PL-TR-93-2012, 1993.

G. C. Sloan, P. C. Tandy, B. E. Pirger, T. M. Hodge, Spatial Structure in the Infrared Spectra of Three Evolved Stars, *BAAS*, 25, 876, 1992.

G. C. Sloan, G. L. Grasdalen, P. D. LeVan, Spatially Resolved Spectra of Silicate Dust Around x Orionis· *Astrophysical Journal*, 404, 328-332, 1993.

G. C. Sloan, G. L. Grasdalen, P. D. LeVan, Spatially Resolved Spectra of the Unidentified Infrared Features Around HD 44179 (The Red Retangle), *Astrophysical Jounal*, 409, 412-416, 1993.

G. C. Sloan, Study of Infrared Emissions from H II Regions, *BAAS*, 24, 1302, 1992.

G. C. Sloan, G. L. Grasdalen, P. D. LeVan, Spatially Resolved Spectra of Silicate Dust Around a Orionis, *ApJ*, 404, 328, 1993a.

G. C. Sloan, G. L. Grasdalen, P. D. LeVan, Spatially Resolved Spectra of the Unidentified Infrared Features Around HD 44179 (The Red Retangle), *ApJ*, 409, 412.

G. C. Sloan, P. C. Tandy, B. E. Pirger, T. M. Hodge, Study of Infrared Emissions H II Regions, *BAAS*, 25, 876, 1993.

G. C. Sloan, Spatially Resolved 10 um Spectra of Circumstellar Material Around Evolved Stars, *BAAS*, 24, 1302, 1992.

G. C. Sloan, G. L. Grasdalen, P. D. LeVan, Study of Infrared Emissions H II Regions, *ApJ*, 404, 328, 1993a, *ApJ*, 409, 412, 1993b.

G. C. Sloan, G. L. Grasdalen, Study of Infrared Emissions H II Regions, in preparation.

T. Su, R. A. Morris, A. A. Viggiano and J. F. Paulson, Internal and translational energy effects on the charge-transfer reaction of CO_2 + with O_2, *J. Chem. Phys.* 94, 8426 1990.

R. Snieder, The Role of the Born Approximation in Nonlinear Inversion, *Inverse Problems*, 6, 247-266, 1990.

J. M. Thomas, F. Kaufman, and M. F. Golde, Energy Transfer Processes Involving Metastable Molecular Nitrogen A State, N_2 (A $^3\Sigma^+$ u, v), *J. Chem. Phys.* 86 1987, 6885.

J. M. Van Doren, S. E. Barlow, C. H. DePuy and V. M. Bierbaum, Chemistry of C_2- and HC_2- with Nitrogen, Oxygen, and Nitrogen Oxides, *Int. J. Mass Spectrom. Ion Processes*, 109, 1991, 305.

Vasco, D. W., Resolution and variance operators of gravity and gravity gradiometry, *Geophysics*, 54, 889-899, 1989.

A. A. Viggiano, R. A. Morris, F. Dale, J. F. Paulson, K. Giles, D. Smith and T. Su, Chemistry of C_2- and HC_2- with Nitrogen, Oxygen, and Nitrogen Oxides, *J. Chem. Phys.*, 93, 1990, 1149.

A. A. Viggiano, R.A. Morris and J. F. Paulson, Chemistry of C_2- and HC_2- with Nitrogen, Oxygen, and Nitrogen Oxides, *J. Chem. Phys.*, 90, 1989, 6811.

A. A. Viggiano, R. A. Morris, J. M. Van Doren and J. F. Paulson, Internal and translational energy effects on the charge-transfer reaction of CO_2 + with O_2, *J. Chem. Phys.* 96, 270 1992.

A. A. Viggiano, J. M. Van Doren, R. A. Morris, J. F. Paulson, Internal and translational energy effects on the charge-transfer reaction of CO_2 + with O_2, *J. Chem. Phys.*, 93, 4761 1990.

A. A. Viggiano, R. A. Morris, J. F. Paulson, Internal and translational energy effects on the charge-transfer reaction of CO_2 + with O_2, *J. Chem. Phys.*, 90, 6811 1989.

A. A. Viggiano, R. A. Morris, J. F. Paulson, Internal and translational energy effects on the charge-transfer reaction of CO_2 + with O_2, *J. Chem. Phys.*, 89, 4848 1990.

A. A. Viggiano, R. A. Morris, F. Dale, J. F. Paulson, K. Giles, D. Smith, T. Su, Internal and translational energy effects on the charge-transfer reaction of CO2 + with 02, *J. Chem. Phys.*, 89, 4848 1990.

L. R. Watson, T. L. Thiem, R. A. Dressler, R. H. Salter, E. Murad, Thermochemistry of Unstable meatl oxides: Techniques, Part I - Position-Sensitive Detection, Part II - Energetic Neutrals Generated By Laser Vaporization of Solid Samples, *J. Chem. Phys.*, 1991, 95, 8944.

L. W. Wolf, D. B. Stone, and J. N. Davies, Crustal Structure of the Active Margin, South Central Alaska: An Interpretation of Seismic Refraction Data From the Trans-Alaska Crustal Transect, J. of Geoph. Res., Vol. 96, No. 96, No. B10, pp.16,455 - 16,469, Sept. 10, 1991.

Wolf, L. W., A. R. Levander, Upper crustal structure of the accreted Chugach terrane, Alaska, *J. Geophys. Res.*, 94, 4457-4466, 1989.

Wolf, L. W., W. K. Wallace, Upper crustal structure of the Chugach and Prince William terranes, Alaska, *Eos Trans.* AGU, 69, 44, 1988.

Wolf, L. W., and D. B. Stone, Upper crustal velocity structure in a convergent margin setting, southcentral, Alaska, *Eos Trans.* AGU, 70, 43, 1989.

Wolf, L. W., A. R. Levander, and G. Fuis, Upper crustal structure of the accreted Chugach terrane, Alaska, *Eos Trans.* AGU, 67, 44, 1195, 1986.

Oral Presentations

Armstrong, P. S., Analysis of Nitric Oxide Observed in Two Field Experiments, CIRRIS 1A and EXCEDE III, Meeting of the Division of Atomic, Molecular, and Optical Physics, May, 1992.

Armstrong, P. S., Analysis of Nitric Oxide Observed in Two Field Experiments, CIRRIS 1A and EXCEDE III, Meeting of the American Geophysical Union, December, 1992.

Armstrong, P. S., Analysis of Nitric Oxide Spectra Taken During the CIRRIS 1A Experiment and on the Constraints Imposed on Possible Production Mechanisms, Meeting of the American Geophysical Union, May, 1993.

Cicerone, R. D., Fracture Characterization From Vertical Seismic Profiling Data, Annual Reservoir Delineation Consortium Meeting at the Earth Resources Laboraory at MIT, October, 1991.

Cicerone, R. D., Fracture Characterization From Vertical Seismic Profiling Data, Meeting of the American Geophysical Union, December, 1991.

Cicerone, R. D., Scattering and Attenuation in New England Derived from Short-Period Seismograms, The Fall AGU Meeting, November 1992.

Dentamaro, A., Calculations for State Specific Electronic Quenching Rates for $^{14}N_2^+$ and $^{15}N_2^+$, The Molecular Collisions Conference, July, 1987.

Dentamaro, A., Calculations for State Specific Electronic Quenching Rates for $^{14}N_2^+$ and $^{15}N_2^+$, The S.P.I.E. Meeting, April, 1988.

Dodd, J. A., Relaxation of Vibrationally Excited OH by O_2, AFOSR Molecular Dynamics Contractors Conference, October, 1989.

Dodd, J. A., Relaxation of Vibrationally Excited OH by O_2, American Geophysical Union Fall Meeting, December 1989.

Eastes, R., s3-4 Data, American Geophysical Union Meeting, April, 1987.

Erickson, G. M., Physical Stability Limit on Sunward-Convection Models, AGU Meeting, December, 1988.

Erickson, G. M., The Reason for Magnetospheric Substorms, Journal of Geophysical Research, April, 1989.

Erickson, G. M., Stability of the Earth's Magnetosphere, Journal of Geophysical Research, April, 1989.

Erickson, G. M., On The Existence of Steady-State Convection in the Magnetotail and the Physics of the Harang Discontinuity, AGU Spring Meeting, May 1989.

Erickson, G. M., Equilibrium Modeling with Anisotropic Pressure, Meeting at Rice University, May, 1989.

Farley, R. W , LIDAR Poster Presentation, CLEO '92, May, 1992.

Foley, J. E., Use of Simulated Annealing Optimization Techniques Applied To Seismic Waveform Data, M.I.T. Earth Resources Laboratory Workshop on Seismic Wave Propagation and Inversion in Heterogeneous Media, July, 1990.

Gardner, II, J. A., Cross Section and Product Time-of-Flight Measurements of the Reaction of N_2^+ with H_2O and D_2O at Suprathermal Energies, Conference on the Dynamics of Molecular Reactions, July, 1989.

Gardner, II, J. A., Cross Section and Product Ion Time-of-Flight Measurements, The Sixteenth International Conference on the Physics of Electronic and Atomic Collisions, July, 1989.

Gardner, II, J. A., Cross Section and Product Time-of-Flight Measurements of the Reaction of N_2^+ with H_2O and D_2O at Suprathermal Energies, Eleventh East Coast ICR and Ion-Molecule Reaction Symposium, October 1989.

Gardner, II, J. A., Reactive Collisions of N^+ and N_2^+ at Suprathermal Energies, 1990 Gordon Research Conference, July, 1990.

Handel, M. D., Subcritical Tropical Cyclone Intensification, Atmospheric and Oceanic Waves and Stability Conference, October, 1991.

Handel, M. D., On the Asymmetry of Tropical Cyclone Outflows, A talk at PL/GPA, April 1992.

Hawkins, R., Simplified Expression for the Total Internal Partition Sum as a Function of Temperature, 43rd Molecular Spectroscopy Symposium at Ohio State University, June, 1989.

Hawkins, R., CO_2 Spectroscopic Constants Determined by Global Least-Squares Fitting and Direct Numerical Diagonalization, 43rd Molecular Spectroscopy Symposium at Ohio State University, June, 1989.

Howard, S. J., Fast Algorithm for Implementing the Minimum-Negativity Constraint for Fourier Spectrum Extrapolation II, Rome Air Development Center (RADC), April, 1988.

Howard, S. J., Fast Algorithm for Implementing the Minimum-Negativity Constraint for Fourier Spectrum Extrapolation II, Optical Society of America (OSA), November, 1988.

Howard, S. J., Fast Algorithm for Implementing the Minimum-Negativity Constraint for Fourier Spectrum Extrapolation II, Optical Society of America in Rochester, October, 1987.

Krause-Polstorff, J., Presented Paper at the International Meeting for Wave Propagation in Random Media & Consultation at Univ. of Washington on Turbulence, August, 1992.

Kuchar, T. A., The Boston University - Arecibo Galactic HI Survey, 181st Meeting of the American Astronomical Society, January, 1993.

Kuchar, T. A., Infrared Properties of HII regions, American Astronomical Society, June, 1993.

Makhlouf, U., Crossed Path Technique and Meteorological Effects on Greenwood Frequency, Laser Guide Star Workshop at Phillips Lab, NM, March 1992.

Makhlouf, U., Intercomparison of Model Predictions for Temperature Inferred from Airglow by Various Techniques, SOAR (Structured Optical Atmospheric Radiance), April, 1992.

Makhlouf U., OH Diurnal Photochemical-dynamical Model, CIRRIS Spectral Data Analysis Meeting, May, 1992.

Makhlouf, U., Is the Krassovsky Ratio Dependent on the Method by Which Temperature is Inferred from the Airglow?, The CEDAR Meeting, June, 1992.

Makhlouf, U., OH Modulation by Gravity Waves, The SOAR Workshop, October, 1992.

Makhlouf, U., Realistic Photochemical-dynamical Hydroxyl Model with a Realistic Background Atmosphere and A Realistic Wind Profile, Chapman Conference, November, 1992.

McCaffrey, R., The Fault Plane Solutions of Earthquakes in Indonesia and New Guinea, IOC/WESTPAC Symposium on Marine Science in the Western Pacific, December, 1986.

McCaffrey, R., The Fault Plane Solutions of Earthquakes in Indonesia and New Guinea, Lamont-Doherty Observatory of Columbia University, March, 1987.

McCaffrey, R., Active Tectonics of the Banda Arc, Indonesia, AGU Meeting, May, 1987.

McCaffrey, R., Depths and Mechanisms of Large Shallow Earthquakes in the Southern Tadjik Basin, Afghanistan, AGU Meeting, April, 1987.

McCaffrey, R., Seismograms from the Indonesian Regional Seismograph Network at the Meteorological and Geophysics Institute, The Fall AGU Meeting, December 1987.

McCaffrey, R., Seismograms from the Indonesian Regional Seismograph Network at the Meteorological and Geophysics Institute, Center for Earthquake Research and Information at Memphis State University, January, 1988.

McCaffrey, R., The Passcal Experiment, Binghamton, and at Boston College, February, 1988.

McCaffrey, R., The Analysis of the Australian Earthquakes, The Symposium on Earthquakes in the Eastern United States at the Univ.of Illinois, April, 1988.

McHugh, M., Solar Neutrino Detection Using Sapphire Crystals, XIIIth Moriond Workshop--Perspectives in Neutrinos, Atomic Physics and Gravitation, Switzerland, February, 1993.

Morris, R. A., Vibrational Quenching of NO+ Ions, Gaseous Electronics Conference (GEC), October, 1987.

Morris, R. A., Vibrational Quenching of NO+ Ions, 41st Annual Gaseous Electronics Conference (GEC), October, 1987.

Nguyen, B. V., The Source Characteristics of the Nuclear Explosion STARWORT, The Americvan Geophysical Union Spring Meeting, May, 1989.

Nguyen, B. V., Source Studies of the FLASK and CALABASH Nuclear Explosion, The American Geophysical Union 1990 Spring Meeting, May, 1990.

Popecki, M., ELF/VLF Waves at the CRRES Satellite, The American Geophysical Union Meeting, December, 1992.

Popecki, M., Alternating Intensity of Electromagnetic and Electrostatic Waves Outside the Plasmaphere as Observed by CRRES, The American Geophysical Union Meeting, May, 1993.

Sloan, G. C., Sources of 13 um Emission, 181st Meeting of the American Astronomical Society, January, 1993.

Stevens-Miller, A. E., Gas-Phase Acidities of Electron Attachment to Transition-Metal Hydrides, University of Delaware, March, 1991.

Stevens-Miller, A. E., The Electron Attachment to $HCo(PF_3)_4$, $HRh(PF_3)_4$, $HMn(CO)_5$, and $HRe(CO)_5$, 201st American Chemical Society Meeting, April, 1991.

Stevens-Miller, A. E., The Acidity and Electron Attachment Work, Division Seminar, April, 1992.

Thomas, J. M., Energy Transfer from Electronically Excited States of Nitrogen, AFOSR Contractors' Meeting, Oct.-Nov., 1988.

Van Doren, J., C_2- and HC_2- Chemistry, and Modeling of the CO_2^++O_2 Reaction, 201st American Chemical Society National Meeting, April, 1991.

Van Doren, J., Internal Energy Effectsin Reactions, SRI International, CA, September, 1991.

Van Doren, J., Does Tickling Help?: Internal Energy Dependences, The Thirteenth East Coast ICR and Ion Molecule Reaction Symposium at Univ. of Delaware, October, 1991.

Van Doren, J., Marine Aerosol Chemistry, PL Community and Harvard University's Chemistry Department, February, 1992.

Vasco, D. W., Linear Inversion of Gravity and Gravity Gradiometry: A Tale of Two Operators, 16th Annual Gravity Gradiometry Conference, February, 1988.

Vasco, D. W., Linear Inversion of Gravity and Gravity Gradiometry, The American Geophysical Union Meeting, May, 1988.

Watson, L. R., Experimentation Palladium/Silicon Mixtures in Various Ratios, American Chemical Society Meeting, April, 1991.

Watson, L. R., Metal Atoms, 12th International Mass Specrometry Conference, September, 1991.

Willemann, R., Deep Focal Mechanisms-Indicators of Stress or Strain Rate, Fall Meeting of the American Geophysical Union, December 1989.

Willemann, R., A Self-Consistent Model of Lithospheric Bending and Stretching During Subduction, Annual Fall Meeting of the American Geophysical Union, December, 1990.

Willemann, R., Stress Propagation in Subducted Lithosphere Modeled as a Thin Elastic Plate, Spring Meeting of the American Geophyscial Union, May, 1990.

Willemann, R., Stress Propogation in Subducted Lithosphere, The Institute for Geophysics, May, 1990.

Willemann, R., Statistical Analyses of Intermediate Depth Seismicity, Annual Meeting of the Seismological Society of America, March, 1991.

Willemann, R., The Global Depth Distribution of Deep Earthquakes, Spring Meeting of the American Geophysical Union, May, 1991.

Willemann, R., Model of Stress Diffusion in Subducted Lithosphere and Temporal Changes in Intermediate Depth Seismicity, National Ocean Service, Maryland, June, 1991.

Willemann, R., Model of Stress Diffusion in Subducted Lithosphere and Temporal Changes in Intermediate Depth Seismicity, National Earthquake Information Center, Colorado, June, 1991.

Willemann, R., Model of Stress Diffusion in Subducted Lithosphere and Temporal Chagnes in Intermediate Depth Seismicity, National Geophysical Data Center, Colorado, June, 1991.

Wolf, L. W., Examples of Data Collected by the USGS and Stanford University, Seismological Society of America, May, 1990.

Wolf, L. W., Seismic Hazard Related Study on Liquefaction, Seismological Society of America, May, 1990.

Wolf, L. W., AFGL Broadband Instrument, Seismological Society of America, May, 1990.

Wolf, L. W., The PACE Experiment, The American Geophysical Union, December, 1990.

Wolf, L. W., Crustal Structure of the Colorado Plateau, GSA Rocky Mountain/Central Meeting, April, 1991.

PROGRAM STATISTICS

<u>Distribution of Scholars by Gender</u>

Total Males	36
Total Females	<u>06</u>
Total Scholars	42

<u>Distribution of Scholars by State</u>

Alaska	1
California	4
Colorado	2
Maryland	1
Massachusetts	18
Mississippi	1
Missouri	1
New Hampshire	2
New Mexico	2
New York	2
Oklahoma	1
Pennsylvania	2
Texas	2
Virginia	1
Washington, DC	1
Wyoming	1

<u>Distribution of Scholars by School</u>

Boston College	4
Boston University	2
California Institute of Tech.	2
Cornell University	2
Florida State University	1
Johns Hopkins University	2
Massachusetts Instiututute of Technology	4
Ohio State University	2
Princeton University	2
Rice University	2
St. Lawrence University	1
St. Louis University	1
Stanford University	1
State Univ. of New York @ Binghamton	1
Univ. of California @ Berkeley	1
Univ. of California @ Santa Cruz	1
Univ. of Cincinnati	1
Univ. of Colorado @ Boulder	5
Univ. of New Hampshire	1
Univ. of New Mexico	1
Univ. of Pittsburgh	1
Univ. of Texas @ Austin	2
Univ. of Wyoming	1
Washington University	1

LIST OF RESEARCH REPORTS
USAF-SCEEE PHILLIPS LABORATORY SCHOLAR PROGRAM

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

CIRRIS 1A AND EXCEDE III MEASUREMENTS OF COLUMN DENSITIES OF ROTATIONALLY EXCITED NITRIC OXIDE

Prepared by: Dr. Peter S. Armstrong

Research Location: Phillips Laboratory,
                                      Geophysics Directorate
                                      Hanscom AFB, MA 01731

AF Research Colleague: William Blumberg

Contract Number: F19628-86-C-0224

CIRRIS 1A and EXCEDE III Measurements of Column Densities of Rotationally Excited Nitric Oxide

Peter S. Armstrong
30 June 1993

Abstract:

Absolute, v-dependent column densities of the thermalized and rotationally excited nitric oxide have been determined from data obtained with the CIRRIS 1A Space Shuttle experiment and the EXCEDE III rocket experiment. The extent of rotational excitation exhibited in the rovibrational fundamental band of NO(X $^2\Pi$) has been investigated for a wide range of atmospheric conditions. In addition, the NO spin-orbit manifold populations have been found to depart from thermal equilibrium, representing a third degree of freedom, along with vibration and rotation, that are not in equilibrium. These results provide important inputs to models of the chemistry, infrared radiance, and energy budget of the thermosphere.

Introduction:

The CIRRIS 1A experiment was flown on board the Space Shuttle during mission STS-39, which was launched on 28 April 1991. The experimental equipment comprised of a Michelson interferometer capable of high resolution (0.63 cm^{-1}), several radiometers, and two photometers. The details of the instruments have been published previously [AHM90,BAR92]. From this experiment, high S/N, earth limb spectra were collected for a variety of tangent heights and atmospheric conditions. For nitric oxide, spectra obtained from tangent heights of 95 km to the orbiter altitude were analyzed by a nonlinear least-squares fit to determine the column densities of NO in the field of view.

The EXCEDE III experiment also used a Michelson interferometer to record the nitric oxide spectra. This interferometer collected spectra using four detectors, covering four separate frequency bands, of which nitric oxide emissions were present in three bands. The interferometer observed an artificial aurora produced by an electron beam. The electron beam produced a large number of

19

excited atomic nitrogen and oxygen in the field of view, which recombine to produce vibrationally and rotationally excited nitric oxide.

The rovibrational fundamental spectrum of nitric oxide has its band center at 1876 cm^{-1} (5.3 μm). Nitric oxide has an unpaired electron, and the ground state is a regular $^2\Pi$ molecule. Therefore, NO has emissions from two spin orbit manifolds, Ω=1/2 and 3/2. Since the ground state of NO is a $^2\Pi$ state, NO rovibrational emissions contain a Q branch (ΔJ=0) in addition to the P (ΔJ=+1) and R (ΔJ=-1) branches.

The most prominent features of the spectrum are the P, Q, and R branches of the v=1 emissions. The daytime spectra show sharp features at 2021 and 1989 cm^{-1}, and these features are v=1 and v=2 R branch band heads, respectively. The presence of band heads in the fundamental rovibrational spectrum indicate population of highly excited rotational states.

The spectral analysis uses a nonlinear least squares fitting routine to determine the rotational distribution for each vibrational level of the radiating NO molecules. The results of the fits are used to determine the tangent height dependencies of the molecular population distributions and the temperature used to describe the thermal distribution.

Experimental:

For both experiments, the analyzed spectra were collected by Michelson interferometers, which were capable of high resolution. The interferometers were referenced with an internal helium-neon (HeNe) laser. For CIRRIS 1A, the detectors were sampled once for each fringe of the laser light. The focal plane of the interferometer consisted of five Si:As detectors. For EXCEDE III, the detectors were sampled either on each HeNe fringe or once every two fringes, depending on the frequency response of the particular detector.

The data acquisition system used on CIRRIS 1A monitored continuously the experimental parameters needed for the complete determination of the solar zenith angle, latitude and longitude of the orbiter and the tangent point, and the mission elapsed time. For EXCEDE III, many

supporting instruments measured the atmospheric conditions, such as visible and ultraviolet spectrometers, photometers, and diagnostics for the electron beam.

For CIRRIS 1A, the interferometer had a filter wheel for limiting the light frequency range passing through the instrument. Interferograms collected using the open filter (2.5 to 20 μm) and filter 3 (4.8 to 12 μm) were transformed to produce the NO fundamental spectrum. The transformation routine applied the post-flight calibration to generate spectra in absolute radiance units. For all NO spectra, triangular apodization was used in the transformation code to reduce the number of sidelobes associated with a single spectral feature and to have an analytic lineshape.

Spectral Fitting Techniques:

The spectral analysis codes are based on codes previously developed to study both laboratory and field spectra of hydroxyl [DOD91, DOD93]. A dual Boltzmann method is used to describe the population of the rotational states, one for the "thermalized" distribution and one for the rotationally excited distribution. The codes use a nonlinear least-squares fitting routine to determine the linewidth and spectral shift of the spectra, and the effective temperature for each distribution. The population of each vibrational level is a free parameter in the fit.

The analysis codes are capable of fitting the populations of the individual spin orbit manifolds for each vibrational level, or of constraining the manifold populations to a ratio determined by the rotational temperature of that distribution. The ability to allow each manifold population to vary is important to all fits to the data, but is most important to the fitting of the high resolution data, especially of the v=1 Q and P branches.

The spectral constants used to determine line positions of NO are based on the work of Goldman [GOL75]. A modification of the centrifugal term, D_v, has been made to accurately position the R branch band heads, which are formed from emissions of highly excited rotational states. The band-averaged Einstein coefficients are from laboratory studies of chemiluminescent NO [RAW89].

Results:

The field observations from the CIRRIS 1A mission produced a large database of infrared spectra. Inside this database is an extensive set of spectra from the fundamental rovibrational emissions of nitric oxide. The set contains global cover-age of NO spectra occurring under nighttime quiescent, daytime, and auroral conditions. While the v=1 airglow component is the prominent feature of the spectra, emissions from highly excited rovibrational states are present in the spectra.

Emissions from highly excited rotational states of NO were observed in both auroral and daytime conditions. The most notable indicator of population in these states is the R branch band head. The states that contribute to a band head also radiate in the far P branch for the particular vibrational level. For v > 1, the column density of the rotational excited component is greater than the thermal component. These conditions constrain the models for the production and destruction of NO.

An additional result from the fits to the data is the ratio of the spin-orbit manifolds is not in equilibrium with the thermal rotational temperature. When the ratio is set to the thermal rotational temperature, the fit systematically overpredicts emissions from the Ω=3/2 manifold and underpredicts the Ω=1/2 manifold emissions. For high resolution spectra, systematics are observed in both the Q branch and the section of the P branch where the emissions from the separate manifolds are separated from each other. For low resolution spectra, the fit to the Q branch provides the only position to check for the spin orbit ratio. The analysis of the spin orbit ratio has been based on the emission from the thermal component. The analysis of the spin orbit ratio provides an additional constraint on the production of NO.

Using the information of the highly excited rotational states and the spin orbit ratio, a systematic analysis of the database commenced to obtain tangent height dependencies of the vibrational column densities. As described in the fitting techniques section, the analysis determined column densities for both thermal and rotationally excited components for each vibrational level.

For daytime conditions, there exists a trend for both the thermal and rotationally excited

components to fall off by a factor of one hundred from tangent heights of 120 km to 250 km. The nighttime profile also shows that the thermal component drops by a factor of one hundred, but the rotationally excited component has more scatter than the equivalent daytime distribution. The radiance levels of the nighttime quiescent spectrum are smaller by a factor of ten compared to a daytime spectrum at the same tangent height. The emissions from the rotationally excited states are just above the noise level for the nighttime scans. Since the R branch band head is the single largest feature indicating that highly excited rotational states are populated, the fit has difficulty in finding the corresponding P branch emissions in the noise.

The results from the EXCEDE III mission shows that the electron-beam excited atmosphere produces a vibrationally and rotationally rich spectrum, with up to ten vibrational levels determined by the fitting routines. The thermal rotational distributions are described by the local kinetic temperature for the altitude that the spectrum was collected. The effective temperature used for the rotationally excited distribution is relatively constant, and a value of 4500 K was used in the final fits to the data.

The results of the fits show that the ratio of rotationally excited to thermal molecules increases with altitude, up to a ratio of one, which occurs at the apogee of the flight (115 km). Correlating the absolute column densities from the fits with the auxillary instrument measurements will be used to constrain the atmospheric models used to predict the amount of nitric oxide under auroral conditions.

References

AHM90   M.A. Ahmadjian, R.M. Nadile, J.O. Wise, and B. Bartschi, J. Spacecraft Rockets, *27*, 669, 1990.

BAR92   B. Bartschi, A. Steed, J. Blakeley, M. Ahmadjian, J. Griffin, and R. Nadile, Proc. SPIE, *1765*, 64, 1992.

DOD91   J.A. Dodd, S.J. Lipson, and W.A.M. Blumberg, J. Chem. Phys., *95*, 5752, 1991.

DOD93   J.A. Dodd, W.A.M. Blumberg, S.J. Lipson, J.R. Lowell, P.S. Armstrong, D.R. Smith, R.M. Nadile, N.B. Wheeler, and E.R. Huppi, Geophys. Res. Lett., *20*, 305, 1993.

GOL75   A. Goldman and S.C. Schmidt, J. Quant. Spectrosc. Radiat. Transfer, *15*, 127, 1975.

RAW89   W.T. Rawlins, M.E. Fraser, and S.M. Miller, J. Phys. Chem., *93*, 1097, 1989.

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

UPPER BOUND ESTIMATES OF ANOMALOUS ION PRODUCTION IN SPACE BASED CRITICAL IONIZATION VELOCITY EXPERIMENTS

Prepared by: Dr. Rodger Biasca

Research Location: Phillips Laboratory,
                                      Geophysics Directorate
                                      Hanscom AFB, MA 01731

AF Research Colleague: Gregory Ginet

Contract Number: F19628-86-C-0224

Upper Bound Estimates of Anomalous Ion Production in Space Based Critical Ionization Velocity Experiments

Rodger Biasca
Geophysics Scholar
Phillips Laboratory, Hanscom Air Force Base, Massachusetts

FINAL REPORT
Contract F19628-86-C-0224

June 25, 1993

Abstract

The critical ionization velocity (CIV) is an anomalous ionization mechanism first proposed by Alfvén. Experiments have confirmed the existence of a critical velocity in laboratory experiments, but sounding rocket experiments have been ambiguous as to the existence of the critical velocity in the ionosphere. The purpose of this paper is to produce upper bound estimates of anomalous ion production in space based experiments of the critical ionization velocity. The analysis relies on the results of implicit particle-in-cell simulations and a simple rate model to predict the number of ions produced as a neutral cloud traverses a point in space. The model assumes a point release of neutral gas in the ionosphere which is meant to represent a typical sounding rocket experiment. The results of the model suggest why strong evidence of CIV is not observed in space based experiments. Space based experiments require the anomalous ionization process to be initiated through seed ionization provided by charge exchange or other mechanisms. This seed ionization process is too slow in space experiments to ignite CIV. The results also indicate that some of the space based results can be accounted for by assuming a large barium-oxygen charge exchange cross section instead of invoking anomalous ionization mechanisms.

1 Introduction

Alfvén [3] has proposed that a neutral gas propagating across a magnetized background plasma will undergo an anomalous ionization when the neutral's drift energy perpendicular to the magnetic field exceeds the ionization energy. This is the critical ionization velocity (CIV) process. The "critical velocity" for the anomalous ionization may be defined as

$$v_c = \sqrt{\frac{2e\phi_{ion}}{m_n}}$$ (1)

where m_n and ϕ_{ion} are the mass and ionization potential of the neutral respectively.

The existence of the critical velocity has been verified in a series of laboratory experiments (see, for example, *Axnäs* [4], *Brenning* [8], *Danielsson,* [9], and *Danielsson and Brenning* [10]) involving a hydrogen plasma drifting across a cloud of neutral helium. The experiments were designed so that the drifting hydrogen would experience a collisionless interaction with the helium neutrals as the clouds collided. Classically the interaction between collisionless gases should be negligible and the hydrogen should be expected to drift through the helium. Instead these experiments showed a strong interaction between the hydrogen and helium. The interaction caused the relative hydrogen-helium velocity to drop to the critical velocity, electrons to heat anisotropically along the magnetic field, and an increase in the plasma density. The drop in the relative hydrogen-helium velocity to the critical velocity is evidence of the CIV process at work.

Because of the success of laboratory experiments, several experiments have also been carried out from sounding rockets and satellites to test for CIV in the ionospheric plasma (e.g. Porcupine [18], Star of Lima [36], Star of Condor [37] CRIT-I [32], and CRIT-II [31]). The results of these space based experiments have in general been inconclusive or negative. Enhanced ionization, hot electrons and enhanced electrostatic fields near the lower hybrid frequency have been observed in some experiments, but other experiments have provided negative evidence of CIV.

The disparity between laboratory and space based experiments has led to further investigation of the basic mechanism driving CIV. This research has included both analytical work (see, for example, *Abe* [1], *Goertz et al.* [17], *Möbius et al.* [28], *Formisano et al.* [13], and *Galeev* [14]), as well as numerical work [2, 23, 24, 25, 27, 29, 6].

This previous work has led to the generally accepted argument that CIV is a process driven by electrostatic instabilities. The general sequence of events is that (1) seed ionization of the neutral background forms an ion beam propagating across a magnetic field, (2) the ion beam is unstable to electrostatic instabilities (e.g. the modified two stream instability (M2SI), (3) the M2SI effectively transfers energy from the ion beam to the electrons, (4) the electrons heat to energies above the ionization energy of the neutrals, (5) further neutrals are ionized by electron impact ionization by the hot electrons, reinforcing the ion beam and leading to a positive feedback loop.

Numerical work using particle-in-cell (PIC) codes has been useful in verifying the nonlinear feedback mechanism of CIV. The use of PIC code simulations however has been restricted by the computational time necessary to complete a simulation. To reduce the times necessary for the simulations PIC simulations have used unphysical mass ratios ($m_i/m_e = 100$) and reduced dimensions. As shown in previous work [6] the restriction of reduced mass ratios may be relaxed by using the direct implicit PIC method [5]. Reduced dimensionality is still required however.

Since implicit PIC techniques allow for the use of realistic mass ratios but current computational limitations still require reduced dimensions, fully consistent models of experiments covering a region of space tens of kilometers in length and time scales on the order of seconds are not yet possible. The numerical models do however provide much insight into the physical processes occuring during CIV and allow for the development of scaling laws and numerical estimates of the times scales and anomalous ionization rates that occur in CIV. Combining the estimates and scaling laws obtained from simulations with a rate model to determine the effects of CIV in an actual experimental release form the basis of this work.

To be more specific, the simulation method used here is a one-dimensional, periodic implicit particle-in-cell code as in previous work[6]. Obviously the use of a one-dimensional code does not allow effects such as the rapidly decreasing neutral density from the gas expansion, losses of hot electrons, and polarization issues to be directly simulating. Instead the simulations are used here to develop scaling laws of the time scales and anomalous ionization rate associated with the CIV process. These time scales and anomalous ionization rate are then used in a rate model to estimate the ion production from CIV during an experimental release.

Since the numerical modeling still requires great simplification over the experimental conditions, the focus throughout the paper will be on producing "upper bound" estimates of ion production by CIV, i.e. estimates of the highest level of ion production which should reasonably be expected from CIV. A one-dimensional code is consistent with producing the "upper bound" estimates that are of interest here. The constant neutral density and lack of escape mechanisms for hot electrons tend to produce the highest anomalous ionization rate and fastest time scales for the CIV process. These in turn lead to the highest production of ions which can reasonably be expected from CIV.

The sections below will first discuss the simulations used to develop estimates of the time scales and ionization rates associated with CIV. This is followed by the application of the results of the simulations in a rate model used to develop upper bound estimates for ion production in CIV. The results of the rate model are compared with the results of several different CIV experiments.

# 2 CIV Simulations of Space Based Experiments

The work begins with a set of simulations designed to provide insight into space experiments involving CIV. As mentioned above the purpose of this section is to develop scaling laws for the time scales and anomalous ionization rates associated with CIV. The simulations consider the gas nitric oxide propagating across an ambient oxygen plasma. The gas NO is chosen since it was recently released during the IBSS CIV experiments on board the space shuttle (STS 39, April 1991). Since the focus here is on developing scaling laws, the precise model used is not critical since other simulations using different gases have shown the results are applicable to other gases and collisional models.

The initial conditions for the current simulation are a quiet, Maxwellian oxygen plasma in a magnetic field. The magnetic field is oriented so that $B_z/B_x\sqrt{m_e/m_i} = 1.0$ and has a strength such that $\Omega_e/\omega_{pe} = 1.0$. The temperature of the plasma is $T_i = T_e = 0.2eV$. A neutral beam with a density $n_n/n_{e0} = 10^6$ and temperature $T_b = 0.2eV$ is propagating in the z-direction across the simulation region with velocity V_n. As in the previous work [6] the neutral beam is not simulated directly. The neutral beam properties are considered a constant over the course of the simulation. The simulations include electron elastic collisions, non-resonant and resonant charge exchange and electron impact ionization. The collisional cross sections are taken from the literature as functions of energy (for example, electron elastic [39], non-resonant charge exchange [7], resonant charge exchange [30], electron impact ionization [21]).

The methodology adopted here is to first present a typical simulation of CIV and to discuss some definitions of characteristic time scales that will be useful in describing the CIV process. After discussing a typical run each portion of the CIV process will be described in turn. Again the emphasis here will be in developing some semi-empirical estimates of the time scales involved in CIV.

2.1 Typical Simulation of CIV

Figure 1 shows typical results of the simulation ($V_n/v_c = 1.5$ where v_c is the critical velocity). The upper figure presents the time history of the electron number and the bottom figure shows the electron temperature as a function of time. As the simulation begins, an ion beam begins to form through charge exchange reactions between the beam and ambient ions. This beam quickly transfers energy to the electrons through the modified
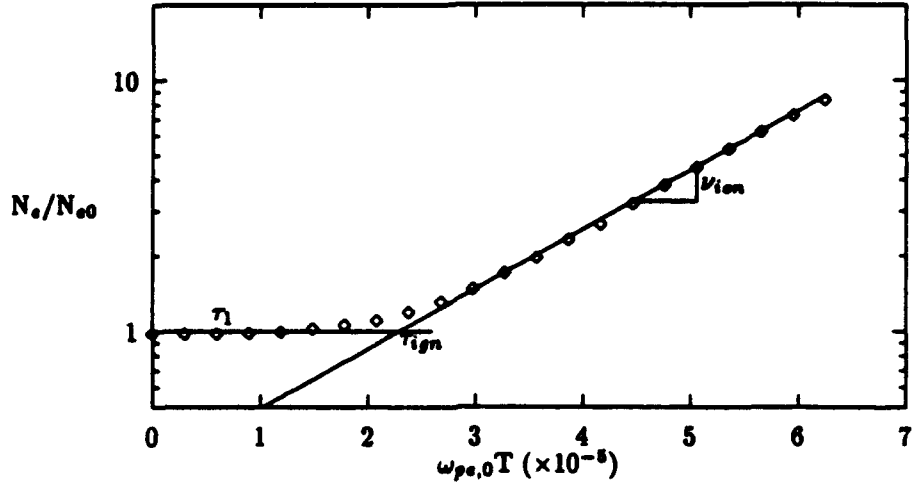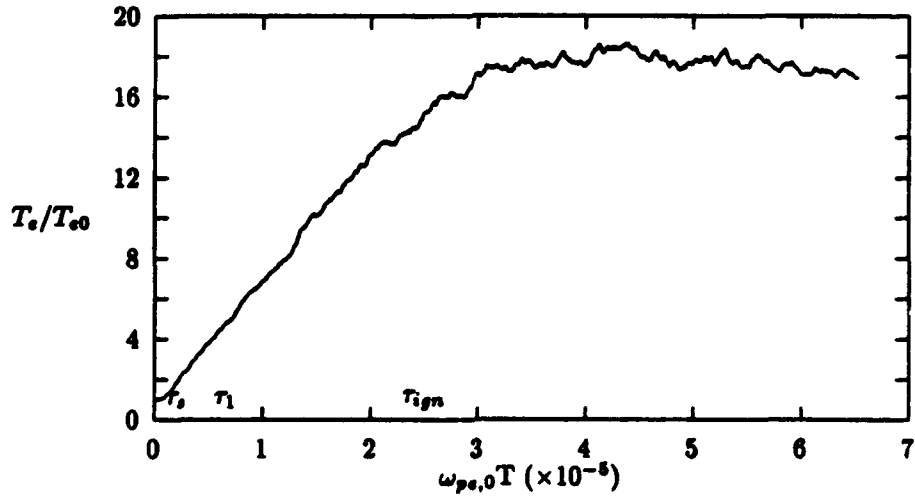
Figure 1: Typical simulation results. (a) Electron number vs time shows an exponential growth. (b) Electron thermal energy vs time displays saturation.

two-stream instability (M2SI). The energy transferred heats the electrons. The process continues with the beam being formed and re-energized through charge exchange. The electrons continue to heat until they reach energies sufficient for ionization. At this point the beam is quickly reinforced through ground ionization while electrons are heated by the M2SI but lose energy to ionization. The electron energy saturates at a point where the energy transferred from the waves is equal to the energy lost to ionization collisions. After saturation the growth in the electron number becomes simply exponential and leads to a definition of the anomalous ionization rate, ν_{ion}:

$$\frac{N_e}{N_{e,0}} = \exp(\nu_{ion} t) \tag{2}$$

The following sections will be concerned with describing the simulations during these processes and developing estimates and scaling laws for the different time scales and anomalous ionization rate. The CIV process can be divided into four regimes: (1) an initial stage from $t = 0$ to τ_s involving the development of the lower hybrid waves, (2) an electron heating stage from τ_s to τ_1, where τ_1 is the time at which first ionization occurs, (3) a transient regime between τ_1 and the ignition time τ_{ign}, where the electron temperature saturates, and (4) an asymptotic state regime in which equation (2) is satisfied. Each of these regimes is discussed in the following sections.

2.2 Initial Stage of CIV

As the simulation starts, charge exchange reactions between the NO neutrals and O^+ ions initiate an ion beam. At this stage the beam is being created at a linear rate of

$$\frac{dn_b}{dt} = \nu_{nrcex} n_{O^+} \tag{3}$$

where ν_{nrcex} is the non-resonant charge exchange rate given by

$$\nu_{nrcex} = <\sigma_{nrcex} v_r> n_{neut} \tag{4}$$

and σ_{nrcex} is the cross section, v_r is the neutral-O^+ relative velocity and n_{neut} is the neutral density.

As the beam begins to build it also begins to transfer energy to the electrons. As the first few ions enter the beam some amount of time, τ_s, will be required to establish the lower hybrid wave spectrum that eventually acts to heat electrons. During this time the energy transfer rate from ions to electrons should be quite low since there is no coherent wave structure to act as the energy transfer mechanism. Linear theory of the M2SI shows that the characteristic time scale for the waves in the M2SI is the lower hybrid period [26]. Hence τ_s can be expected to be on the order of a lower hybrid period, where the lower hybrid frequency is given by

$$\omega_{LH} = \omega_{pi}(1 + \omega_{pe}^2/\Omega_e^2)^{-1/2}. \tag{5}$$

2.3 Electron Heating Stage

After the initial establishment of the wave spectrum the ions begin to transfer energy to the electrons. At least during the very initial stages this energy transfer process can also be expected to be linear in time. The total electron kinetic energy can be found by assuming that the fraction of energy transferred from the newly created ions to the electrons is a constant, and that the time to transfer energy from the beam to electrons is small compared to the time scale of interest (e.g. $\omega_{LH}^{-1} \ll \Omega_i^{-1}$). Since the number of newly born beam ions is simply given by $n_{O^+}\nu_{nrcex}t$ the electron kinetic energy is given by

$$n_e(T_e - T_{e0}) = \nu_{nrcex} n_{O^+}(t - t_0)\eta\frac{1}{2}m_n V_n^2 \tag{6}$$

30

where η is the fraction of energy transferred from ions to electrons and $t_0 = \mathcal{O}(\omega_{LH}^{-1})$. A discussion of the possible values of η can be found in *Formisano et al.* [13] and were discussed in previous work [6]. Those discussions however focused on determining η in the limit where CIV is ignited and the electron temperature is saturated. Under these conditions Formisano correctly ignores the original plasma ions because their density becomes increasingly negligible with respect to the exponentially growing beam ion density, and determines that if the ions are effectively unmagnetized ($\nu_{ion}/\Omega_i \gg 1$) $\eta = 2/3$, and $\eta = 0.025$ for highly magnetized ions. Previous numerical work [6] has confirmed the results of Formisano in the same saturation limit. In the current case however at early times when the background ion distribution dominates, η remains high even for time scales long compared to the ion gyroperiod. The results here indicate that for this initial heating stage η can be expected to be initially high (≈ 0.5) and assuming it to remain high for the duration of this period is consistent with upper bound estimates.

To verify the above comments Figure 2 shows the results for the initial stages of the simulation. The upper figure shows the beam ion density while the lower figure shows the kinetic energy as a function of time. As seen in the upper figure the ion beam density increases linearly in agreement with the remarks made above. The solid line on the figure is the value predicted by the charge exchange rate using $v_r = V_n$. The actual number of beam ions is somewhat above the value predicted by the charge exchange rate due to statistical deviations since the number of particles created is still low.

The electron kinetic energy starts with a region of no detectable heating. This constant electron temperature lasts on the order of $5000\omega_{pe}^{-1}$ which is consistent with the local lower hybrid frequency of the beam ($\omega_{LH}^{-1} \approx 2000\omega_{pe}^{-1}$) assuming $n_b/n_e = .01$). After about $5000\omega_{pe}^{-1}$ the electron temperature begins to increase roughly linearly. The rate of increase of the electron temperature is consistent with a constant efficiency of $\eta = .54$.

2.4 First Ionization

The start of ionization occurs at τ_1 when the energetic tail of the electron distribution reaches the ionization energy, $e\phi_{ion}$, of the neutrals. Although the exact electron energy distribution depends on many details of the wave particle interaction heating the electrons, it is assumed here that the distribution can be characterized by a bulk electron temperature that is equal to some fraction, f, of the ionization energy when ionization begins, i.e.

$$T_{e\phi} = f e\phi_{ion} \tag{7}$$

Along with equation (6) this observation gives the time required for the tip to reach $e\phi_{ion}$. After the tip of the distribution function reaches the ionization energy of the neutrals ionization will not be observed for some period after this related to the ionization frequency, i.e. $\tau_{ion} = \nu_{ion}^{-1}$. Since the electrons will continue to heat during this short time of τ_{ion}, at the start of ionization the electron bulk temperature will be given by

$$T_{e,ion} = T_{e\phi} + \tau_{ion}\nu_{ces}\eta\frac{1}{2}m_n V_n^2. \tag{8}$$

assuming the efficiency remains constant and the temperature is increasing linearly. Now, τ_1 may be found by combining equation (6) and (8) and taking $n_e \approx n_i$,

$$\tau_1 = \tau_s + \tau_{ion} + \frac{T_{e\phi} - T_0}{\nu_{nrces}\eta\frac{1}{2}m_n V_n^2} \tag{9}$$

The efficacy of equation (9) is indicated in Figure 3 which compares the results from the equation with the results of the simulations. The model parameters derived from the comparison are included in the figure. The efficiency η is about 50%, the constant $\Omega_i(\tau_s + \tau_{ion})$ is approximately 0.16, and $f = T_{e\phi}/e\phi_{ion} = .10$.
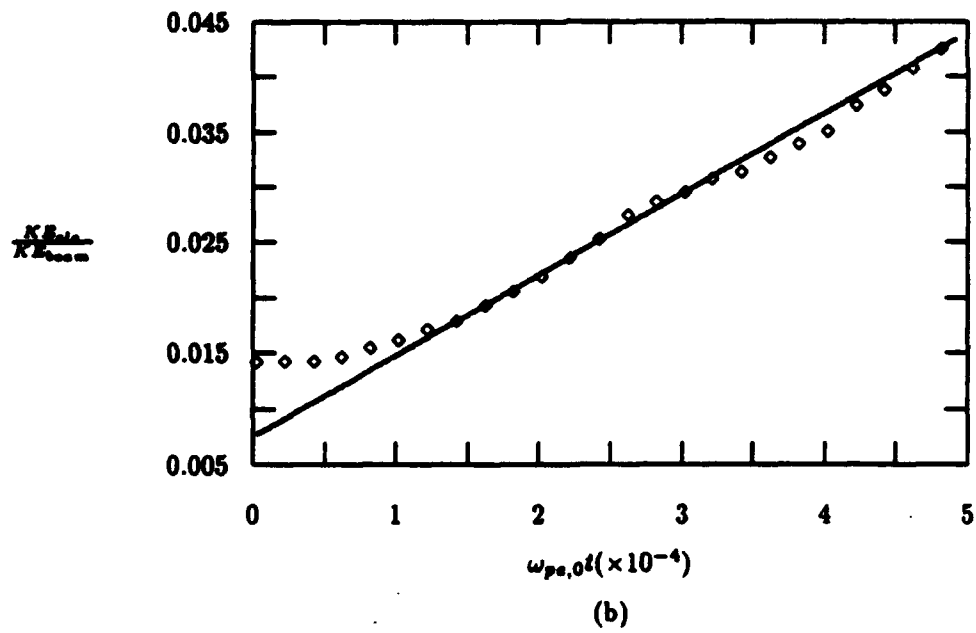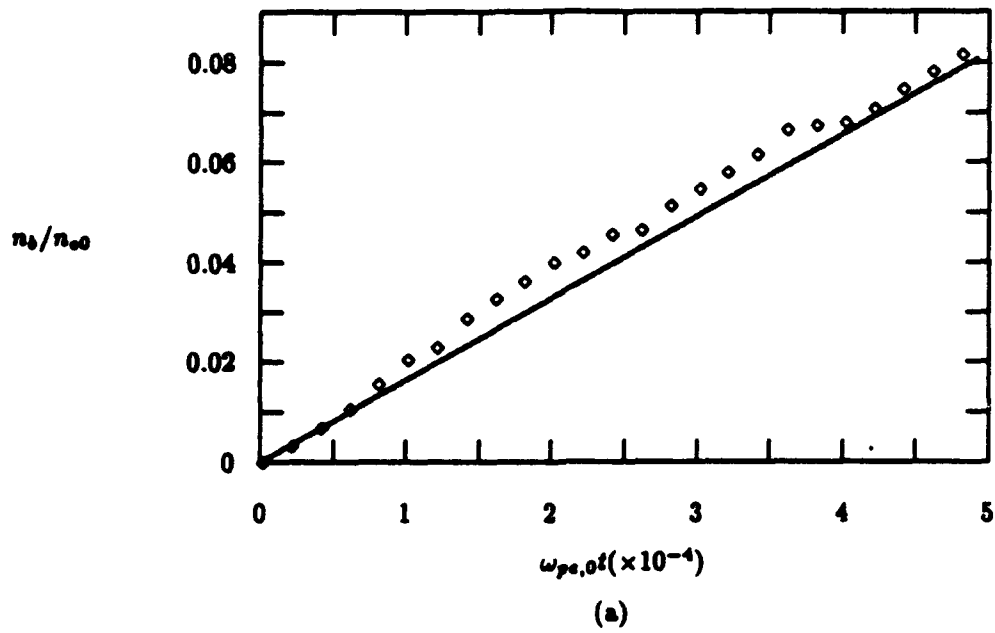
Figure 2: Initial stages of CIV. (a) Beam density. During the initial stage of CIV, charge exchange reactions increase the beam density linearly with time. (b) Electron kinetic energy. Since the efficiency of energy transfer to the beam is constant, the electron kinetic energy also increases linearly.
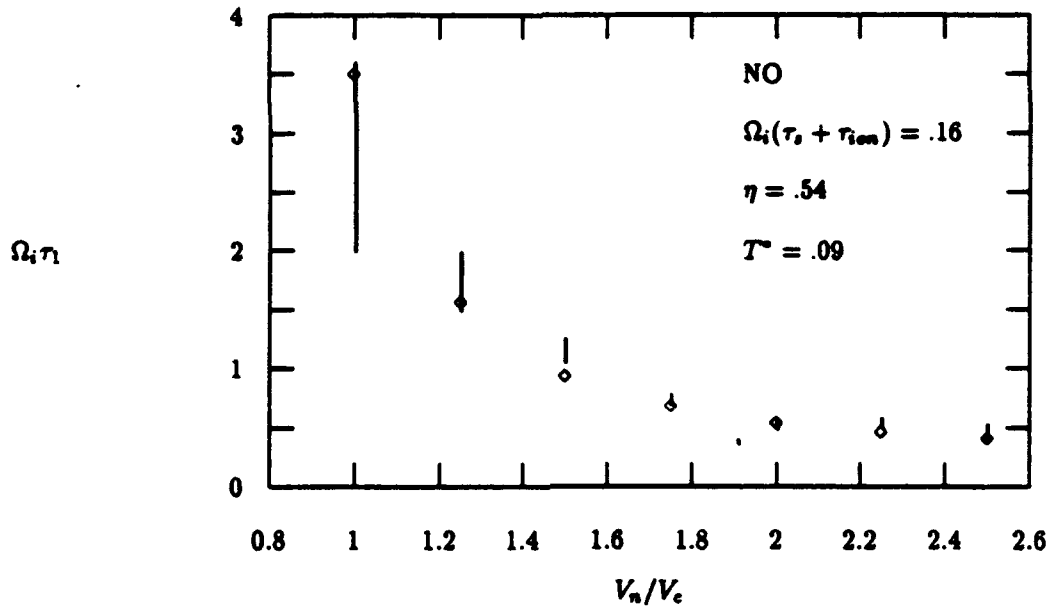
Figure 3: Time to first ionization. The vertical bars represent the results of the simulations, the markers are the prediction of the equation (9). For the simulation results, the values plotted represent the range between the time between the first electron impact ionization event and the third electron impact ionization event
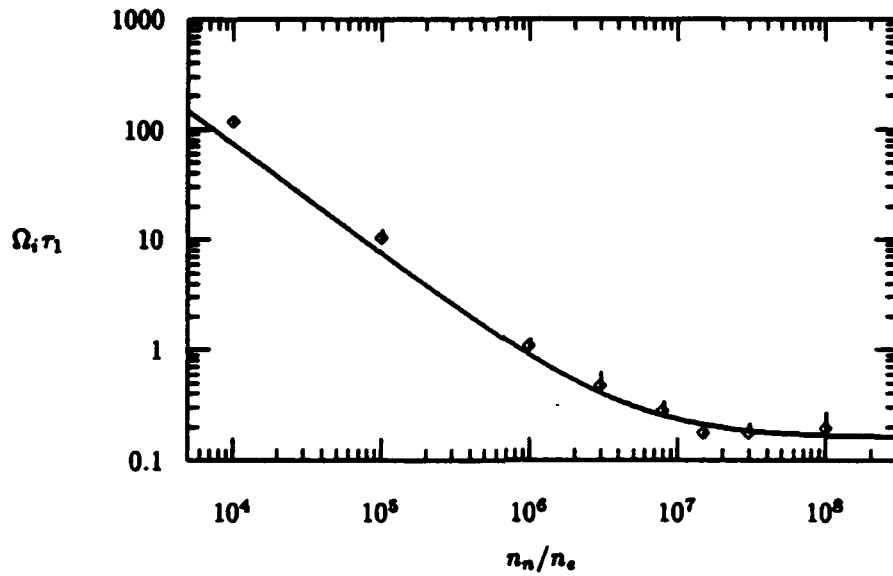

Figure 4: Dependence of the time for first ionization on the neutral density. The results agree well over four orders of magnitude.

Simulations with other gases and other initial conditions indicate that these values are insensitive to the precise simulation values.

Figure 4 shows similar results as a function of the neutral density. The parameters are the same as used in Figure 3. Good agreement is seen over four orders of magnitude.

2.5 Transition Phase

After the first ionization events the electron kinetic energy will continue to increase towards a saturation value. An estimate of the time needed to enter the exponential phase can be obtained by defining the "ignition time" as the time needed to reach the break point as illustrated in Figure 1. The ignition time is a function of the detailed collisional processes and energy gain/loss mechanisms occurring after ionization begins. In this regime the linear temperature increase assumption used in the previous section is no longer valid (the simulation shown in Figure 1 indicates the temperature continues to climb linearly well beyond τ_1; other simulations indicate this is not universally true, the linear phase of temperature increase often ends near τ_1). A specific expression for the ignition time is difficult to obtain however an estimate can be obtained by considering the ratio of the ignition time to the time for first ionization, τ_{ign}/τ_1. For most cases the ignition time is roughly 3 to 5 times the time for first ionization. This estimate is roughly true as a function of both velocity and density. Later, in considering the rate model of CIV, this transition regime will be ignored by setting $\tau_{ign} = \tau_1$. This is again consistent with looking for upper bound estimates of ion production by CIV.

2.6 Exponential Growth State

The electrons continue to heat until energy losses to ionization and excitation are equal to energy added from wave transfer. Once the exponential growth state is reached the plasma density increases as

$$\frac{dn_e}{dt} = \alpha_{ion}\nu'_{ion}n_e \tag{10}$$

where α_{ion} is the fraction of the electron population with energies above the ionization energy, and ν'_{ion} is just a redefinition of the anomalous ionization rate, $\nu'_{ion} = \nu_{ion}/\alpha_{ion}$. The introduction of α_{ion} and ν'_{ion} is a simple acknowledgment of the fact that only electrons with energies above the ionization energy of the neutrals are capable of ionizing. The precise magnitude of the anomalous ionization rate, ν_{ion}, is difficult to estimate a priori because it is sensitive to the details of the electron distribution function. However the scaling of ν_{ion} with the neutral velocity and density can be suggested through simple arguments. The redefined anomalous ionization rate can be written as

$$\nu'_{ion} = <\sigma_{ion}v> n_n \tag{11}$$

where the angle brackets represent an average taken over the portion of the distribution function with $E > e\phi_{ion}$. Over the energy ranges of interest here the quantity $<\sigma_{ion}v>$ is assumed to be constant.

The remaining portion of the anomalous ionization rate, α_{ion}, can be found by considering the idealized saturated electron distribution function as given by *Formisano et al.* [13] and used subsequently by *Lai et al.* [22] and others. In the idealized case the hot tail of the electron distribution function simply has a constant density from below the ionization energy to the beam energy. For this distribution function the fraction of electrons with energies above the ionization energy is simply proportional to the beam energy. Combining this with equation (11) gives the scaling law for the anomalous ionization rate,
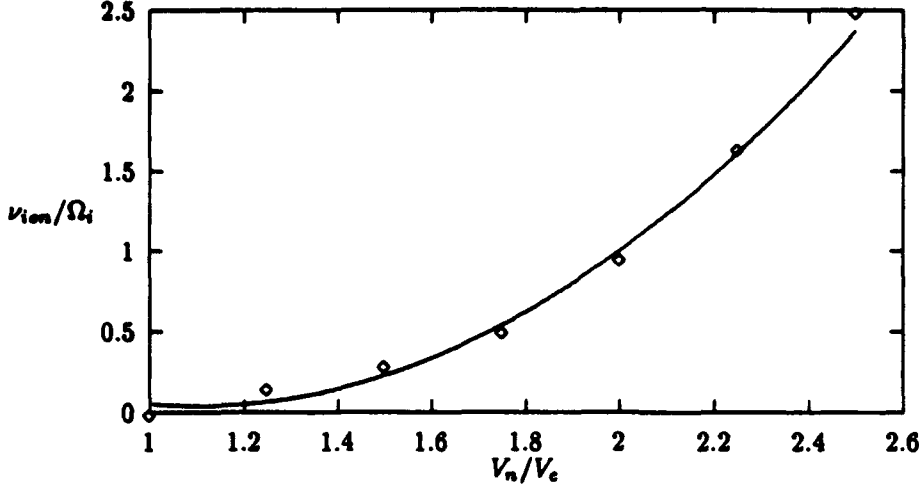
$$\nu_{ion} \sim n_n(V_n^2 - V_c^2)/V_c^2 \tag{12}$$

34

Figure 5: Anomalous ionization rate vs neutral beam velocity for NO. Below $V_n/V_c = 1$, $\nu_{ion}/\Omega_i \approx 0$. The anomalous ionization rate increases quickly as the neutral velocity exceeds the critical velocity.

Although derived through very simple arguments the simulations as described below confirm the efficacy of this result.

The anomalous ionization of NO is first considered as a function of the neutral beam velocity. For these runs the neutral density is chosen such that $N_n/N_{e,0} = 10^6$. Figure 5 shows ν_{ion}/Ω_i after electron temperature saturation for NO as a function of the neutral velocity (V_n/V_c). The markers shown in the figure represent data points from the simulation results. The solid lines are a fit to the data points. As seen in the figure, for $V_n/V_c < 1$, $\nu_{ion}/\Omega_i \approx 0$. This is expected since below the critical velocity no anomalous ionization should be observed. Above the critical velocity however the anomalous ionization rate quickly increases. The excellent agreement between the fit and the data indicates the ionization rate's dependence on the energy of the neutral beam.

Also of interest is the effect of the neutral density on the ionization rate. Figure 6 shows ν_{ion}/Ω_i as a function of the neutral density, N_n/N_e. In this figure the neutral beam velocity was held constant at $V_n/V_c = 1.5$ while the neutral beam density was varied.

The ionization rate is seen to scale linearly with density up to a density of $N_{neut}/N_{e0} \approx 10^7$. That is, for $N_{neut}/N_{e0} < 10^7$

$$\frac{\nu_{ion}}{\Omega_i} \sim \left(\frac{N_n}{N_{e,0}}\right) \tag{13}$$

Above $N_{neut}/N_{e0} \approx 10^7$ the ionization rate begins to level off. At these high densities the growth in the ionization rate decreases because elastic collisions destroy the hot tail of the electron distribution function. This corresponds to a transition between the "collective" and "resistive" forms of CIV discussed by Machida and Goertz [23].

Finally, although scaling laws have been obtained for the anomalous ionization rate, obtaining an expression for the absolute magnitude is much more difficult. The values shown in Figure 5 were obtained directly by measuring the slope from a semilog plot of the electron density as a function of time after the electron temperature saturated. These values tend to indicate values of the anomalous ionization reaction rate, $\nu_{ion} = \mathcal{O}(\Omega_i)$. Unlike the scaling laws derived above, however, these values may be sensitive to the actual inputs into the simulation. To account for this the rate model below will consider a wide range of
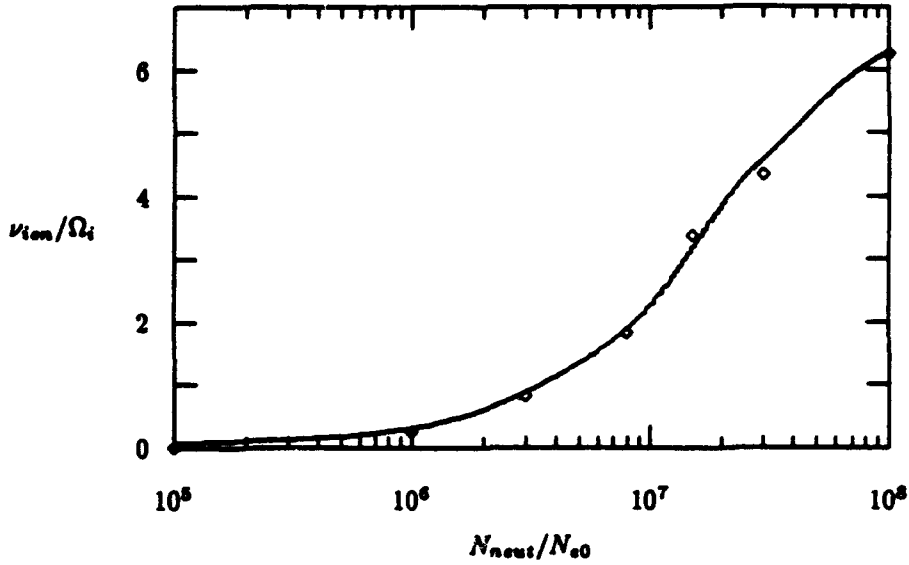
Figure 6: Effect of Density on the anomalous ionization. For the lower values of density, $\nu_{ion}/\Omega_i \sim N_n/N_e$. At higher densities, the transition from collective to resistive becomes evident. (The solid line represents a spline fit to the data)

possible values of ν_{ion} centered around the values measured from the simulations. It should be kept in mind however that although the scaling laws for ν_{ion} can be derived, the actual magnitudes of ν_{ion} are based on empirical measurements from the simulations.

3 A Simple Rate Model

The simulation results provide estimates of time scales needed to initiate CIV as well as scaling laws for ν_{ion}. In addition, although a specific formulation for ν_{ion} has not been developed, the simulation results suggest quantitative estimates of $\nu_{ion}/\Omega_i = \mathcal{O}(1)$. Since full three-dimensional PIC code simulations representing space experiments are not yet computationally possible, this section now uses these results in a rate model to estimate the yield and spatial extent of ion production in space releases. As before, the focus here is on upper bound estimates of the production of ions.

The analysis begins by considering the convection of a neutral gas across a magnetic field. Once a sufficient time has passed to heat electrons to the ionization energy the neutral gas will begin to undergo ionization at the anomalous ionization rate, ν_{ion}. Based on the results of the above simulations the anomalous ionization rate will be assumed to scale linearly with the neutral density. With this assumption the anomalous ionization rate may be expressed as the product of a reaction rate, p_0, and the neutral density, i.e.

$$\nu_{ion} = p_0 n_n \tag{14}$$

where p_0 is independent of the neutral density. The undetermined factor in the equations is the reaction rate, p_0. This reaction rate is taken from the simulation results.

Newly created ion–electron pairs, created at a rate $p_0 n_n n_e$, are assumed to remain on the the magnetic field line on which they were created. Ignoring electron losses along the field is clearly consistent with looking for the maximum creation rate. It is less clear what the effect is of ignoring plasma transport across the field. There is both observation and theory [19, 15, 16, 12] to suggest that when a drifting plasma stream

36

is "born" with sufficient density relative to a background plasma, a polarization electric field will develop and the newly born plasma will continue to drift rather than braking in the ambient plasma rest frame. This electric field is created because (in the absence of full polarization) the cycloid motion of the new ions and electrons displaces them in opposite $\pm \mathbf{v} \times \mathbf{B}$ directions, creating charge layers on the stream boundaries and an electric field that reduces the cycloid separation and induces a drift in the plasma. In the limit of a dense "new" stream covering a large area, the net drift velocity will just be that of the center of mass of the mixed plasma. For smaller less dense streams, the net drift and polarization field strength is difficult to estimate since it is subject to dissipation of the charge layers into the ambient which effectively couples a greater ambient plasma mass into the mix. For the scenario considered here it is postulated that if significant polarization occurs during the charge exchange heating period the effect would be to suppress CIV because heated electrons would be convected to regions of lower neutral density where (as will be shown below) CIV is less likely. Ignoring the effect is therefore consistent with seeking upper bounds. If polarization occurs after ignition of CIV, the effect would again be limiting since if the net drift with the background is limited, the free energy available to the plasma instability will also be limited. Further, it is at least self-consistent to ignore the effect here since it is precluded by the geometry of the simulations. With these assumptions, the plasma density at a point then may be assumed to increase as a result of the ionization process, i.e.

$$\frac{\partial n_e}{\partial t} = p_0 n_n n_e \tag{15}$$

Again, the assumptions leading to equation (15) are consistent with the approach of producing upper bound estimates for the occurrence of CIV.

Solving the above equations requires three steps: (1) determining the neutral density as a function of time for given initial and boundary conditions, (2) finding when the ignition time criterion is satisfied, and (3) solving for the electron density as a function of r and t using the anomalous ionization rates determined in the previous section.

4 Neutral Models

The macroscopic properties of the neutral gas may be found from taking proper moments of the distribution function, f_n. The governing equation for f_n is the Vlasov equation, hence the macroscopic properties of the neutrals may be found from a solution of

$$\frac{\partial f_n}{\partial t} + \mathbf{v_n} \cdot \nabla f_n = \left(\frac{\partial f_n}{\partial t}\right)_c \tag{16}$$

where $\mathbf{v}_n$ is the neutral velocity and $(\partial f_n/\partial t)_c$ represents a source/sink term due to ionization of the neutrals. At least for the initial stages of CIV, a reasonable assumption is that only a small fraction of the neutrals will be ionized. With this assumption, the right hand side of equation (16) can be ignored giving simply

$$\frac{\partial f_n}{\partial t} + \mathbf{v_n} \cdot \nabla f_n = 0 \tag{17}$$

The approximation to the shaped charge velocity distribution used here is

$$f(v) = \begin{cases} 1/v_h & 0 < v < v_h \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

where v_h is some maximum velocity. This simple velocity distribution is chosen to be a reasonable representation of typical [18] shaped charge releases as well as to make the integrations in the following analysis

tractable. A comparison with the actual velocity distribution of typical space releases shows that the flat distribution function may underestimate the number of ions moving at above the critical velocity. This may slightly violate the "upper bound" assumptions which have been built into the model elsewhere. However the results indicate that the dominate factor affecting CIV comes from the density drop associated with the neutrals' expansion into a spherical solid angle and not the precise form of the velocity distribution, per se. Therefore, considering the analytic simplifications, the use of the somewhat idealized distribution function shown in equation (18) seems to be justified.

An initial condition is also needed. The initial condition chosen here is that the neutrals are initially located between r_0 and r_1 and exhibit a $1/r^2$ dependence on the radial distance, i.e.

$$n_n(r, t = 0) = \begin{cases} C_0/r^2 & r_0 < r < r_1 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

In order to match typical experimental conditions, C_0 is chosen as 4×10^{22} m^{-1} to achieve $n_n = 10^{16}$ m^{-3} at $r_1 = 200$ m. The choice of 10^{16} m^{-3} is related to the cutoff due to the windowing effect as discussed in previous work[6], while the $r_1 = 200$ m length is consistent with the start of ionization in the Porcupine data. The lower length of $r_0 = 56$ m is chosen to provide 10^{24} neutral particles within the initial volume assuming a conic half-angle of $\Theta = 13.5°$. The velocity v_h is taken as 13 km/s. With these initial conditions the neutral density as a function of position and time can be found by solving equation (17) for $f(r, v, t)$ and integrating over velocity space. The result is

$$n_n(r > r_1, t) = \begin{cases} \frac{C_0}{r^2} \frac{r_1' - r}{v_h t} & \max(r_1, r_0') < r < r_1' \\ \frac{C_0}{r^2} \frac{r_1 - r_0}{v_h t} & r_1 < r < \max(r_1, r_0') \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where $r_1' = r_1 + v_h t$ and $r_0' = r_0 + v_h t$. This expression is simply a geometric expansion into the spherical solid angle $1/r^2$, with a correction term $(r_1 - r_0)/(v_h t)$ to account for the velocity distribution. A pictorial version of the neutral density at several different times is shown in Figure 7.

5 Ignition Time Analysis

As shown in the simulations, for the space based experiments, the time to start CIV (i.e. the ignition time) is about $3\tau_1$ to $5\tau_1$, where τ_1 is given by equation (9).

In this section the time to first ionization, τ_1, is determined for the neutral gas releases and compared to the transit time, T_{tr}, of the neutral gas past a given element of plasma. For CIV to ignite the transit time of the neutral gas must exceed the time to first ionization, i.e.

$$\tau_1 < T_{tr} \tag{21}$$

The simulations were all performed at a given neutral velocity and density. However, as the results for the neutral model indicate, the density and velocity at a point in space will vary in time. In order to expand the results of the simulations to the current situation a simple integral representation is adopted: the electron energy at a point in space is assumed to follow

$$T_e = T_{e0} + \int \eta \frac{1}{2} m_n V_n^2 < \sigma_{nrcex} v > n_n dt \tag{22}$$

over the linear regime. The term $< \sigma_{ncrex} v >$ is introduced here as the rate for $Ba - O^+$ charge exchange reactions. At this point however the model might be considered somewhat more general by considering the
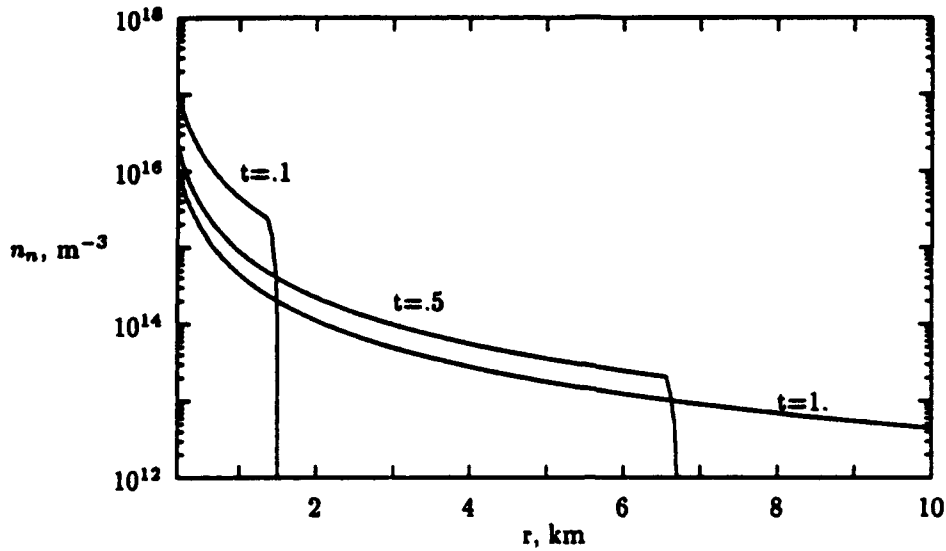
Figure 7: Model Neutral Density for Porcupine. The neutral cloud expands outward from the release point. The neutral density quickly drops during the expansion.

$< \sigma_{nrces} v >$ term to represent the production of barium ions not only from charge exchange but from all non-CIV processes including, for example, the stripping reaction $Ba + O \rightarrow Ba^+ + O + e$. Although reactions other than charge exchange have been considered as the seed ionization mechanism for CIV (*Lai et al.* [22]), it seems most likely that the charge exchange reaction dominates the non-CIV production of ions. Because charge exchange is likely the dominant seed ionization mechanism, the rest of this work will continue to include in the analysis only charge exchange for the seed ionization process. This is also consistent with the simulations which did not include seed ionization mechanisms other than charge exchange.

The time to first ionization is given by τ_1 where τ_1 satisfies

$$T_{e\phi} - T_{e0} = \int_{t_0}^{t_0 + \tau_1} \eta \frac{1}{2} m_n V_n^2 < \sigma_{nrces} v > n_n dt \tag{23}$$

and t_0 is the time at which neutrals first arrive at the point (the small constant $\tau_s + \tau_{ion} \sim \mathcal{O}(\omega_{LH})$ has been ignored here). Equation (23) is somewhat more clearly written as

$$T_{e\phi}^* - T_{e0}^* = \eta \left(\frac{V_n}{V_c}\right)_{max}^2 < \sigma_{nrces} v > \int_{t_0}^{t_0 + \tau_1} n_n dt \tag{24}$$

where $T^* = T/(m_n V_c^2/2) = T/(e\phi_{ion})$, and (V_n/V_c) are evaluated at their maximum values (again consistent with looking for the upper bound for the occurrence of CIV). In addition, the reaction rate, $< \sigma_{nrces} v >$, is assumed to be a constant.

The integral of equation (24) is simplified by taking $(V_n/V_c)_{max} = (v_h \sin\theta/V_c)$, where θ is the angle of the neutral release with respect to the magnetic field. Most experimental releases have been performed with the velocity vector of the neutral jet directed at an angle to the magnetic field. The inclusion of θ in the definition of $(V_n/V_c)_{max}$ is necessary to include this effect in the model and simply states that the component of the neutral velocity perpendicular to the magnetic field is important for CIV, not the full velocity, v_h.

39

Rearranging equation (24) results in

$$\xi \equiv \frac{r_0 v_h}{C_0} \frac{T_{e\phi}^\circ - T_{e0}^\circ}{\eta \left(\frac{V}{V_e}\right)^2_{max} < \sigma_{nrces} v >} = \frac{r_0 v_h}{C_0} \int_{t_e}^{t^\circ = t_e + r_1} n_n dt \qquad (25)$$

where $t_0 = (r - r_1)/v_h$. Physically, the parameter ξ represents the (nondimensionalized) characteristic time necessary to heat electrons to energies sufficient to begin ionization. The parameter ξ is also convenient in that it contains several parameters (e.g. η, C_0, $T_{e\phi}^\circ - T_{e0}^\circ$) which either contain some uncertainty or vary for different experimental conditions. When the results are presented for various values of ξ, the emphasis will be on the variability of the cross section σ_{nrces}. It should be remembered however that different values of ξ may also be thought of as corresponding to varying values of the other parameters included in the definition.

The integral appearing in equation (25) will reappear throughout the work. Performing the integral results in the definition

$$G(r > r_1, t_e, t_b) \equiv \int_{t_e}^{t_b} n_n dt = \begin{cases} \frac{C_\phi}{r^2}\left[\frac{r_1 - r}{v_h}\log\left(\frac{t_b}{t_e}\right) + t_b - t_e\right] & t_0 \le t_e < t_b \le t' \\ \frac{C_\phi}{r^2}\left[\frac{r_1 - r}{v_h}\log\left(\frac{t'}{t_e}\right) + \frac{r_1 - r_0}{v_h}\log\left(\frac{t_b}{t'}\right) + t' - t_e\right] & t_0 \le t_e < t' \le t_b \\ \frac{C_\phi}{r^2}\frac{r_1 - r}{v_h}\log\left(\frac{t_b}{t_e}\right) & t_0 < t' \le t_e < t_b \end{cases} \qquad (26)$$

where $t' \equiv (r - r_0)/v_h$.

Using the definition for G, equation (25) results in

$$\xi = \frac{r_0 v_h}{C_0} G(r, t_0, t^\circ) \qquad (27)$$

The solution to equation (27) can be found through a simple iterative procedure: calculate t° using the second condition of equation (26); if the criterion $t^\circ \ge t'$ is violated, calculate t° from the first condition of equation (26) (which requires a Newton-Raphson iteration or other root finder since this equation is now transcendental).

Once the time to first ionization is known the criterion for CIV to start is that the transit time, T_{tr}, of the neutrals through a plasma element be longer than the time to first ionization, $\tau_1 < T_{tr}$. The transit time can be considered as the time between the arrival at a point in space of the first neutrals and the arrival of the last neutrals still moving at the critical velocity, i.e.

$$T_{tr} = \frac{r - r_0}{v_c / \sin\theta} - \frac{r - r_1}{v_h} \qquad (28)$$

The desire is to compare the results of equation (27) with the results of equation (28) for parameters similar to those of experimental releases. The simulations suggest reasonable numbers which may be used to estimate the time for CIV to start in barium. The values used here are $T_{e\phi} = .1e\phi = .5$ eV, $T_{e0} = .2$ eV, and $\eta = .5$. In addition, as in Porcupine, $v_c = 2.7$ km/s, $v_h = 13$ km/s, and $\theta = 28°$.

Figure 8 compares the time to first ionization, τ_1, and the transit time, T_{tr}. Since the value of the barium-oxygen charge exchange cross section is not accurately known, the solution is shown for a series of values of $< \sigma_{nrces} v >$. The cross section used varies from 10^{-17} cm^2 to 10^{-14} cm^2 in order to cover the range of cross sections assumed by various authors (in the following, all cross sections will be specified at an energy of 8eV). The cross section reaction rates, p_{nrces}, therefore vary from 3.3×10^{-12} cm^3/s to 3.3×10^{-9} cm^3/s. The corresponding values of ξ ($\approx 10^{-4}$ to 10^{-1}, respectively) are also included on the figure. The exponentially
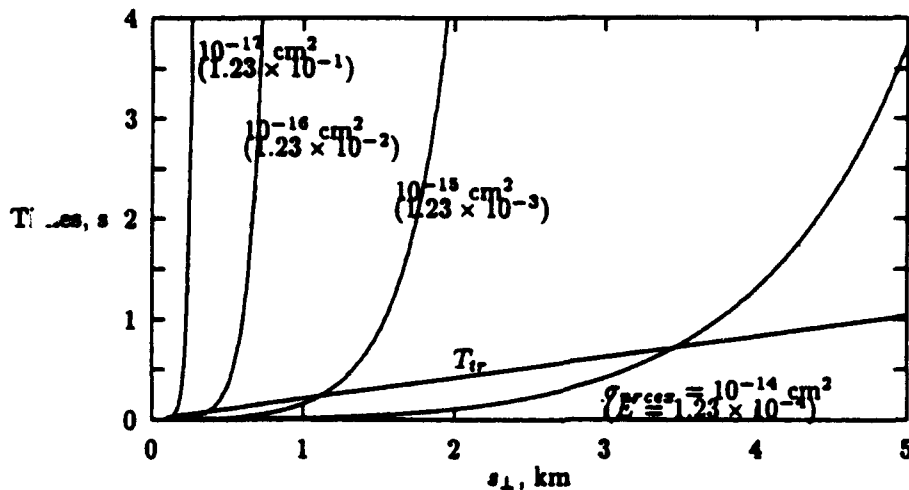
Figure 8: Ignition time and transit time for "Porcupine" release. The exponential-like curves are the time to reach first ionization. The linear curve is the transit time. For CIV to ignite, the ignition time must be less than the transit time.

increasing curves shown on the figure are the time to reach first ionization for the stated value of ξ, as given by the solution of equation (27). The times are shown as a function of the down range distance from the release point ($s_\perp = r\sin\theta$). The linearly increasing line is the transit time as a function of $s_\perp$ (equation 28). As shown in the figure the criterion for CIV to ignite ($T_{tr} > \tau_1$) is only satisfied within several hundred meters of the release point for $\sigma_{nrces} = 10^{-17}$ cm^2, while CIV ignites out to a horizontal range of about 3.5 km for $\sigma_{nrces} = 10^{-14}$ cm^2. This is an important result in relation to the Porcupine experiment. The results of Porcupine indicate enhanced ionization out to $s_\perp$ =15 km from the release point. Even with the largest charge exchange cross section assumed here ($\sigma_{nrces} \approx 10^{-14}$ cm^2), CIV cannot be achieved beyond 3.5 km. The results here indicate that CIV should be much more localized than the results of Porcupine indicate. One factor not considered in the original analysis of the Porcupine data was the charge exchange contamination of the results. *Swenson et al.*[34] have suggested that the charge exchange cross section may be as large as 10^{-14} cm^2. In the following section the possibility that much of the observed ionization during experimental releases was due to charge exchange reactions will also be considered.

6 Exponential Growth Phase

The previous section determined when first ionization would occur for the different neutral models. In this section the exponential growth phase of CIV is examined, and the fractional increase in the plasma density is determined.

Ionization begins at the time $t_0 + \tau_1$. Consistent with the strategy of producing an upper bound, the gas is assumed to enter its exponential phase immediately upon reaching τ_1, i.e. the transition phase described in the section describing the simulation results is ignored. Equation (15) gives the fractional increase in the

plasma density as

$$\frac{n_e}{n_{e0}} = \exp\left[\int_{t_0+\tau_1}^{t_0+T_{ir}} p_0 n_n dt\right] = \exp\left[p_0 G(r, t_0 + \tau_1, t_0 + T_{ir})\right] \tag{29}$$

The results for n_n, τ_1 and T_{ir} from the previous sections can be used to solve this equation for the fractional increase in plasma density.

For the Porcupine release the density model is given by equation (20). τ_1 and T_{ir} are found from equations (27) and (28) respectively. The plot of n_e/n_{e0} as a function of $s_\perp$ shown in Figure 9 is informative. The figure shows the results for various values of the electron heating rate, ξ, and the anomalous ionization reaction rate, p_0, in order to account for uncertainties in the analysis. The figure shows two values of ξ, 10^{-3} and 10^{-4}, corresponding to charge exchange cross sections of approximately 10^{-15} cm^2 and 10^{-14} cm^2 respectively. Some uncertainty also exists in the anomalous reaction rate, p_0, since barium has not been simulated directly and some controversy exists as to the cross sections. The results of the previous section do suggest however that the anomalous ionization rate should be in the range of $p_0 = 10^{-15}$ to 10^{-16} m^3/s. These actual values may however depend more heavily on the specific inputs into the simulations than the scaling laws used elsewhere in the analysis. To account for the possibility of substantially higher or lower p_0's, the results in the figure shown range from $p_0 = 10^{-17}$ to $p_0 = 10^{-14}$ m^3/s.

In agreement with the ignition time analysis the results here show that CIV does not exist except within several kilometers of the release point. In fact, although first ionization will be achieved within 3.5 km for the $\xi = 10^{-4}$ case, substantial increases in the plasma density due to CIV are only observed within about 3 km. Similar results are obtained for other values of ξ: increases in plasma density attributable to CIV are confined to within at most several kilometers of the release point even for the largest anomalous reaction rates.

Considering the results of the simulations and the "upper bound" assumptions that have been built into the current model the most likely anomalous reaction rate to be achieved in an actual experimental release will probably be $\approx 10^{-16}$ m^3/s. As seen in Figure 9, for an anomalous reaction rate $p_0 \approx 10^{-16}$ m^3/s, a peak increase of about 30% in the plasma density is achieved near the release point. The enhanced plasma density also decreases to negligible values within several hundred meters of the release point. For even lower anomalous reaction rates only a small increase in plasma density is seen even near the release point.

These results again have important implications for sounding rocket releases. Because of the decreasing neutral density as the neutral cloud expands from the release point, the time necessary to heat electrons to begin ionization, and the magnitude of the reaction rates likely to be achieved, CIV will not be an efficient ion production mechanism in point releases. Instead, fractional plasma density increases of $\approx 30\%$ may occur within the first several hundred meters of the release point. Beyond several hundred meters the ionization rate due to CIV is negligible.

The relation of these results to the actual experimental evidence obtained from Porcupine is discussed below.

7 A Comparison of "Porcupine" Results

Up to this point, the electron density as a function of the radial distance from the release point has been determined. In this section these results are related to the experimental measurements obtained during the Porcupine release.

In Porcupine two main results were obtained: (1) An estimated 10% of the total neutral population was ionized within 9 sec of the release, and (2) a densitometer tracing taken above the terminator gave the relative intensities.

The densitometer measurements are an indication of the total number of ions appearing above the terminator. The results in the previous section focused on the electron density as a function of radial
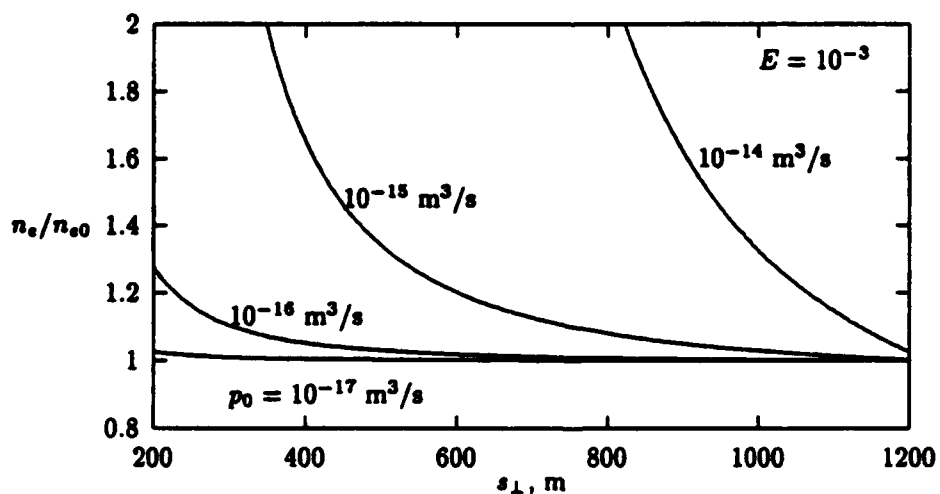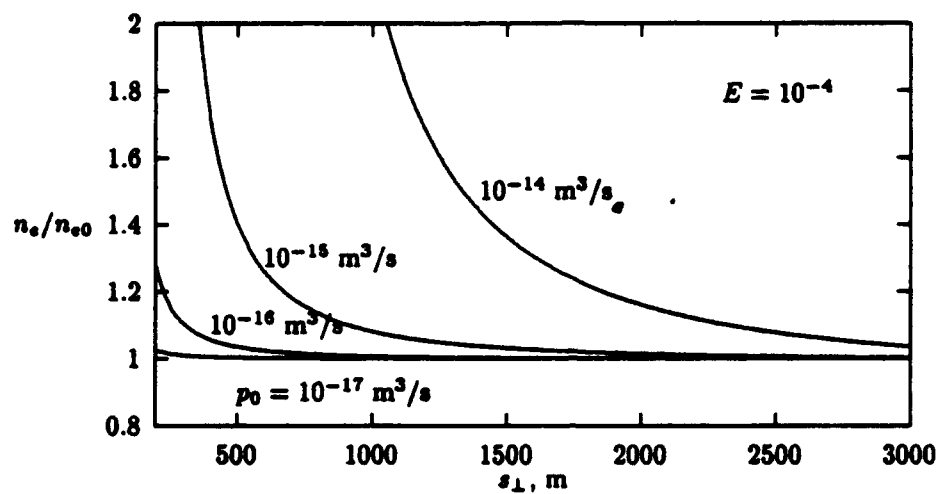
42

Figure 9: Fractional Ionization for Porcupine. The plasma density is seen to be enhanced by CIV only within several kilometers of the release point. For $p_0 = 10^{-16}$ m^3/s, the plasma density is increased by a maximum of only 30%.

distance. For purposes here the number of beam ions appearing above the terminator is assumed to be related to the number of ions produced at a radial position, r, integrated over the spherical volume and projected to a perpendicular distance:

$$I(s_\perp / \sin\theta) = I(r) = 2\pi(1 - \cos\Theta)r^2 n_b \tag{30}$$

In addition, the total yield, Y, of the release may be found by integrating the number of barium ions produced over all of space and normalizing by the initial number of neutral barium atoms, i.e.

$$Y = \frac{2\pi}{N_b}(1 - \cos\Theta) \int_0^\infty r^2 n_b dr \tag{31}$$

where $N_b = 10^{24}$ is the number of initial neutral barium particles.

The ion beam density, n_b, has contributions from CIV and charge exchange, i.e.

$$n_b = (n_b)_{ces} + (n_b)_{CIV} \tag{32}$$

where the CIV component is determined from equation (29) using $(n_b)_{CIV} = n_e - n_{e0}$, and the charge exchange component can be found from solving

$$\frac{\partial(n_b)_{ces}}{\partial t} = \nu_{nrces} n_{O+} = p_{nrces} n_n n_{O+} \tag{33}$$

The remaining parameter needed in order to calculate equations (30) and (31) is the oxygen ion density, n_{O+}. For the determination of the time to first ionization the variation in the ambient ion density was ignored. On the time scale of seconds of interest here the results show this may no longer be a good assumption. Because of the high barium neutral density and possibly large charge exchange cross section, substantial depletion in the ambient oxygen number density may occur near the release point. In experimental releases, of course, the actual level of depletion is determined by the rate of depletion, the flux of ambient oxygen from more distant regions into the depletion region, development of electrostatic forces which may enhance the "refilling" of the depletion region, and other factors. These details of the depletion are beyond the scope of the current analysis. Instead, two limits are considered. In the first limit the depletion of the the oxygen will be ignored, i.e. n_{O+} will be taken as a constant. In the second limit the oxygen will be allowed to deplete, but the "refilling" of the depletion region will be ignored, i.e. the oxygen density in the depletion region will be allowed to go to zero. The actual experimental results should be somewhere between these two limits.

For the first limit, the background oxygen density is assumed to be a constant. In this case equation (32) becomes

$$n_b = p_{nrces} n_{O+} \int_{t_0}^t n_n dt + n_e - n_{e0} = p_{nrces} n_{O+} G(r, t_0, t) + n_e - n_{e0} \tag{34}$$

In the second limit, the ion beam density at a point is related to the electron density and ambient ion densities through

$$n_b + n_{O+} = n_e \tag{35}$$

where the ambient ion density is calculated from

$$\frac{\partial n_{O+}}{\partial t} = -\nu_{nrces} n_{O+} \tag{36}$$

giving

$$\frac{n_{O+}}{n_{O+,0}} = \exp\left[-\int_{t_0}^t p_{nrces} n_n dt\right] = \exp\left[-p_{nrces} G(r, t_0, t)\right] \tag{37}$$

44

The intensity and yield as a function of $s_\perp$ can be calculated from equations (30) and (31). These last steps are carried out numerically, and the results of I vs. $s_\perp$ are presented in Figures 7 and 7 for several different assumptions on the cross sections. Here, it can also be shown that depletion can be ignored for the purpose of calculating the ignition time, τ_1. At τ_1, $G(r, t_0, t)$ may be eliminated in equation (37) using equation (27). Evaluation of equation (37) yields less than 4% depletion at τ_1 for the assumed parameters.

Figure 7 shows the model results at 9 sec assuming a non-resonant charge exchange cross section of 10^{-14} cm^2 for Ba-O$^+$ ($\xi \approx 10^{-4}$). The top panel is a linear x-axis, while the lower panel is just the top panel replotted with a log x-axis to emphasis the effect of CIV near the release point. Results are included for three different values of the anomalous ionization reaction rate ($p_0 = 0.$ (no CIV), 10^{-16} m^3/s and 10^{-15} m^3/s), with and without ambient oxygen ion depletion.

The figure clearly shows that the effect of CIV is small at the assumed values of anomalous ionization reaction rates. Little difference is seen between the results with and without CIV except quite near the release point. In these results nearly all newly formed ions occur through charge exchange; the CIV portion is not significant except within perhaps several hundred meters of the release point. Also included in the figure are markers representing the results obtained by Haerendel during Porcupine. The shape of the relative intensity curve is seen to agree well with the Porcupine results except near the release point. Near the release point the experimental results are bracket by the depletion/no depletion limits. It should be expected that a more complete model of ion depletion in this region would increase the agreement between the model and experimental results.

The given value of total ionization for Porcupine was 20% of the neutrals with velocity in excess of the critical velocity. Equivalently, this represents roughly 10% of the total ion population. The current results assuming $\xi = 10^{-4}$ ($\sigma_{nres} = 10^{-14}$) and $p_0 = 0$ gives 1.3% ionization within 15 km of the release, assuming a background density of $n_{e0} = 2 \times 10^5$/cm^3.

Figure 7 shows similar results, but assuming $\xi = 10^{-3}$. Here, a difference does appear between the results with and without CIV, but this difference is only significant within approximately 1 km of the release point. According to this result, which corresponds more closely to the charge exchange cross section assumed by Haerendel, the results of Porcupine should have produced a more intense streak within 1km of the release point than was actually observed. In addition, this model produces only .13% ionization.

The good agreement between the calculated intensity profiles and those observed for the Porcupine release for $s_\perp > 1$ km is strong evidence for ion production by charge exchange alone. This conclusion is further strengthened by demonstrating that if CIV occurs, the CIV-produced ions should be confined to a more limited region near the release point than actually observed in the Porcupine release. There still exists however the discrepancy between the calculated yield of ions and the ions observed in Porcupine. At this point, this discrepancy must be attributed to either (1) an ambient plasma density that was an order of magnitude higher than the quoted value, (2) a $Ba - O^+$ charge exchange cross section even larger than the maximum assumed value, $\mathcal{O}(10^{-14})$, (3) errors in the photometric determination of yield, or (4) extensive cross field transport of the CIV ions. Only the fourth option admits the possibility of CIV.

8   Comparison with other releases

The results discussed so far have been applied to Porcupine. In this section the implications of these results for the other barium and strontium releases are discussed. More specifically, this section discusses the hypothesis that Porcupine did not observe strong CIV over the entire 15 km range, but instead that much of the observed ionization was due to charge exchange. The effect on charge exchange is consistent with many of the other CIV results observed in space based experiments. In fact, by considering the results of CIV experiments, a probable range for the $Ba - O^+$ charge exchange cross section can be developed.

45
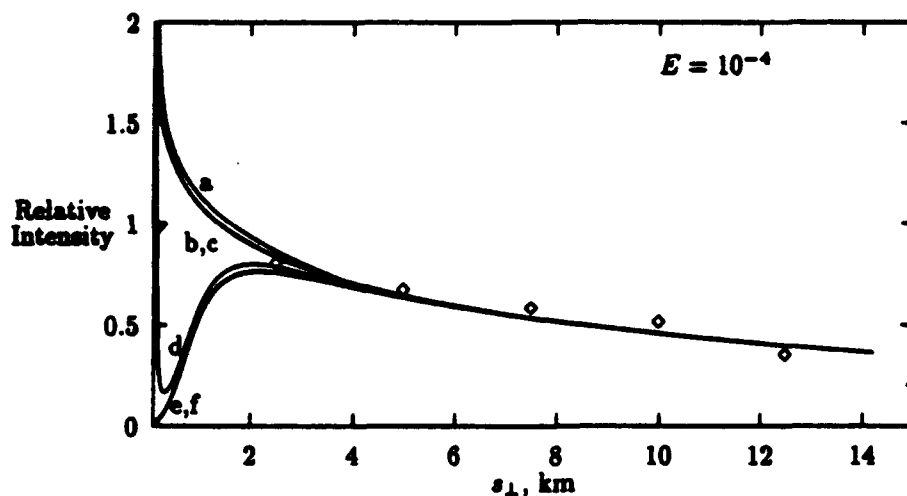
(a)
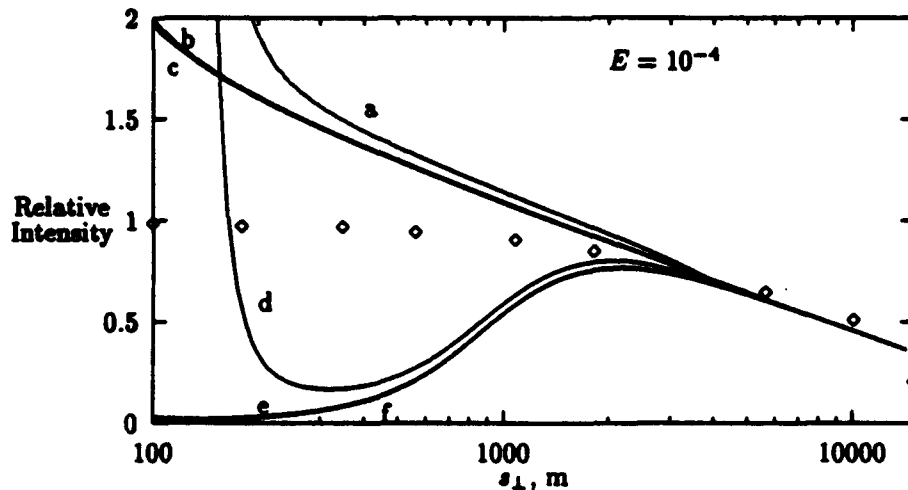
(b)

Figure 10: Comparison of Charge Exchange and Anomalous Ionization in Porcupine. The curves represent the relative intensity calculated from the current results ($\xi = 10^{-4}$). The markers are the data of Porcupine. In this case, the contribution to ionization from CIV is almost insignificant except for the highest $p_0 = 10^{-15} m^3/s$. Even with $p_0 = 10^{-15} m^3/s$, ionization due to CIV is significant only within about 1 km of the release point. (a) $p_0 = 10^{-15}$ m^3/s, without depletion; (b) $p_0 = 10^{-16}$ m^3/s, without depletion; (c) $p_0 = 0.$ m^3/s, without depletion; (d) $p_0 = 10^{-15}$ m^3/s, with depletion; (e) $p_0 = 10^{-16}$ m^3/s, with depletion; (f) $p_0 = 0.$ m^3/s, with depletion; The total calculated yield is 1.3%.

46
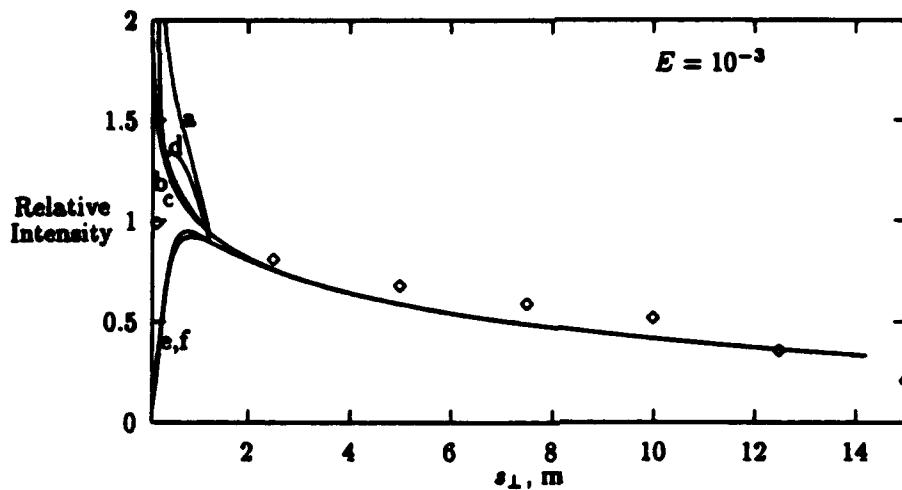
(a)
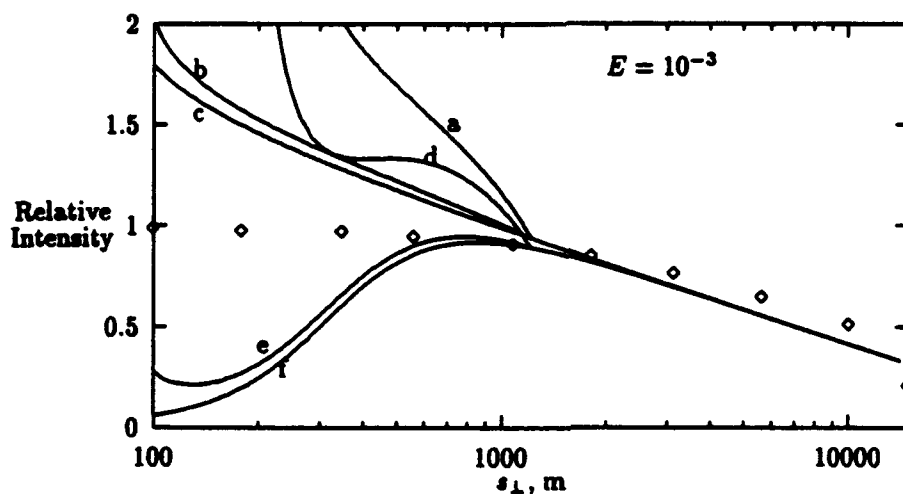
(b)

Figure 11: Same as Figure 10, but with $E = 10^{-3}$. This case results in .13% ionization of the neutral barium. (a) $p_0 = 10^{-15}$ m^3/s, without depletion; (b) $p_0 = 10^{-16}$ m^3/s, without depletion; (c) $p_0 = 0.$ m^3/s, without depletion; (d) $p_0 = 10^{-15}$ m^3/s, with depletion; (e) $p_0 = 10^{-16}$ m^3/s, with depletion; (f) $p_0 = 0.$ m^3/s, with depletion;

8.1 Bubble Machine:

Bubble Machine was a charge release performed in full sunlight consisting of both Ba and Sr [11]. The Ba was expected to photoionize while the Sr was observed for evidence of CIV. Bubble Machine observed up to 50% ionization of the Sr. The error in this figure may however be up to a factor of three. Unfortunately the large uncertainty in the experimental results makes it difficult to judge this release. The observed ionization of the Sr however is consistent with the conclusions drawn for Porcupine. In Bubble Machine the photoionizing Ba would act as a rapid seed ionization mechanism. This seed ionization through photoionization of Ba would likely lead to a more rapid electron heating than observed in Porcupine. The increase in electron heating would decrease the time to first ionization of the Sr and ultimately lead to enhanced values of Sr ionization. Hence "pumping" of the CIV process through photoionization of Ba likely led to the enhanced ionization of Sr observed in this experiment.

8.2 Star of Condor, Sr90:

Star of Condor was a radial strontium release [37]. The release produced Sr moving at all angles to the magnetic field. Since Sr does not photoionize on the time scales of concern for these experiments, the $Sr - O^+$ charge exchange cross section is believed to be no larger than 10^{-16} cm^2, and roughly 50% of the released Sr was estimated to have velocities in excess of the critical velocity, this release was expected to provide an excellent example of the operation of CIV in the ionosphere. The results however indicate virtually no ionization occurred.

The Sr90 release was a conical Sr release occurring at an angle of 45° to the magnetic field [38]. A faint field aligned ion streak was detected, but it could be accounted for through photoionization. If ions produced by CIV were present, the maximum yield was estimated to be 0.18% of the Sr. Like Star of Condor, Sr90 failed to provide evidence of the existence of CIV.

The negative results of these two Sr releases is not unexpected in light of the current work. The small $Sr - O^+$ charge exchange cross section makes these releases substantially different from the Ba releases. Virtually no charge exchange contamination should be expected in the Sr releases. Without CIV no significant ionization should be observed, and indeed minimal ionization was observed in these experiments.

8.3 Star of Lima:

Star of Lima was a barium release carried out from a sounding rocket that overperformed [20, 36, 37]. The overperforming rocket exposed the released barium to more sunlight than expected. The results of the experiment indicate that approximately 5×10^{20} of the 10^{24} released neutrals (.05%) were ionized. *Torbert and Newell* [36] point out that this level of ionization can be accounted for through photoionization in this experiment. However, using a charge exchange cross section 7×10^{-15} cm^2 at 8eV, the measured ambient plasma density of 2×10^4 cm^{-3} and other parameters similar to Porcupine, the current model predicts about .05% of the neutrals would be expected to ionize through charge exchange. Hence, assuming a large cross section, charge exchange reactions can account for the number of ions produced in Star of Lima.

The relative roles of the photoionization and charge exchange mechanisms in producing the ions is not clear from the experimental data available in Star of Lima. It is clear however that CIV does not need to be invoked in order to account for the ionization observed.

8.4 CRIT-I:

The CRIT-I experiment consisted of two conical barium releases performed approximately 45 km below the solar terminator and directed upward toward sunlight [32]. The angle between the axis of the jet and the magnetic field was 45°. In CRIT-I, diffuse ionization was observed several 10's of km's away from the

release point. The analysis of the experimental data led to determination of a time constant for ionization of 1800 sec . In their analysis of the experiment *Stenbaek-Nielsen et al.* [32] assumed a barium-oxygen charge cross section of 10^{-16} cm^2, giving a time constant for charge exchange of 70000 sec. However, using the assumed value of 10^{-14} cm^2, the time constant for charge exchange is 1400 sec. Clearly, if the large Ba-O$^+$ charge exchange cross section is correct charge exchange can account for all the ionization observed in CRIT-I.

8.5 CRIT-II:

Like CRIT-I, CRIT-II also consisted of two barium shaped charges [33, 31, 35]. Similar results were obtained in each release. The releases were performed 100 km below the terminator at an angle of 58° to the magnetic field [31]. CRIT-II observed diffuse ion production at a calculated rate of .8%/sec. In addition, substantial fluxes of high energy electrons [35] and intense electric field near the barium lower hybrid frequency [20] were both observed from in-situ diagnostics.

Originally, using an analysis similar to that used for CRIT-I, ions produced by charge were not considered a significant ionization source. Subsequently, charge exchange contamination of CRIT-II has been examined by *Swenson et al.* [34]. Using a value of 10^{-14} cm^2, charge exchange is shown by Swenson et al. to account for the entire .8%/sec ionization.

8.6 Other Observations:

The predictions here that space releases of Ba will lead to substantial charge exchange reactions, but not CIV are also consistent with the in situ measurements of hot electrons and lower hybrid waves made during Porcupine[18] and Star of Lima [20, 35]. According to the theories developed here the charge exchange reactions should form an unstable beam which will act to heat electrons through the M2SI. In the case of sounding rocket releases the time scales are such that explosive growth in the plasma density is not achieved in CIV, yet the lower hybrid waves of the M2SI and the hot electrons should still be observed. This is indeed the case in Porcupine and Star of Lima.

Finally, a review of the space releases indicates some empirical correlation between increased ambient plasma density and CIV-like processes (enhanced ionization, hot electrons and lower hybrid waves). The releases observing the highest levels of enhanced ionization (Porcupine and CRIT-II) were performed at the highest ambient densities [35]. The hypothesis that charge exchange reactions are the source of the observed enhanced ionization is also consistent with this empirical correlation. Higher ambient plasma densities will lead to an increased rate of charge exchange reactions.

9 Summary

As the comparison above shows, the results of CIV experiments, with the possible exception of Porcupine, indicate the $Ba - O^+$ charge exchange cross section at 8eV may be bracketed between 7×10^{-15} and 10^{-14} cm^2. When compared to the Porcupine results, using these large cross sections and the model developed here, the spatial extent of ionization observed in Porcupine may be explained by non-CIV processes although the yield produced by the current model still predicts substantially less ionization than reported in Porcupine. If the high yield reported in Porcupine was caused by a CIV event, and is not attributable to measurement errors, the models used here indicate that extensive polarization and cross field transport is required to match the observations. Without this effect CIV ionization must be confined to within at most several kilometers of the release point.

49

References

[1] Abe, T. Theory for the Critical Ionization Velocity Phenomenon. *Planet. Space Sci.*, *32*, 903, 1984.

[2] Abe, T. and S. Machida. Production of High-energy Electrons caused by Counterstreaming Ion Beams in an External Magnetic Field. *Phys. Fluids*, *28*, 1178, 1985.

[3] Alfvén, H. *On the Origin of the Solar System.* Oxford University Press, 1954.

[4] Axnas, I. Experimental Investigation of the Critical Ionization Velocity in Gas Mixtures. *Astrophys. Space Sci.*, *55*, 139, 1978.

[5] Barnes, D.C., T. Kamimura, J.N. Leboeuf, and T. Tajima, Implicit particle simulation of magnetized plasma, *J. Comp. Phys.*, *52*, 480, 1983.

[6] Biasca R., D. Hastings, and D. Cooke. Simulation of the Critical Ionization Velocity: Effect of Using Physically Correct Mass Ratios. *J. Geophys. Res.*, *97*, 6451, 1992.

[7] Biondi M. Drift Tube Determinations of Ion-Molecule Reaction Rates at Low Energies. *Comm. Atomic Mol. Phys.*, *6*, 159, 1977.

[8] Brenning, N. Experiments on the Critical Ionization Velocity Interaction in Weak Magnetic Fields. *Plasma Phys.*, *23*, 967, 1981.

[9] Danielsson, L. Experiment on the Interaction between a Plasma and a Neutral Gas. *Phys. Fluids*, *13*, 2288, 1970.

[10] Danielsson, L. and N. Brenning. Experiment on the Interaction between a Plasma and a Neutral Gas, II. *Phys. Fluids*, *18*, 661, 1975.

[11] Deehr, C., E. Wescott, E Stenbaek-Nielson, and H. Foppl. A Critical Velocity Interaction between Fast Barium and Strontium Atoms and the Terrestrial Ionospheric Plasma. *Geophys. Res. Lett.*, *9*, 195, 1982.

[12] Dressler, R., J. Gardner, D. Cooke, and E. Murad. Analysis of Ion Densities in the Vicinity of Space Vehicles: Ion-Neutral Chemical Kinetics. *J. Geophys. Res.*, *96*, 13795, 1991.

[13] Formisano, V., A. Galeev, and R. Sagdeev. The Role of the Critical Ionization Velocity Phenomena in the Production of Inner Coma Cometary Plasma. *Planet. Space Sci.*, *30*, 491, 1982.

[14] Galeev, A. Weak Turbulence Theory of an Enhanced Gas Ionization By the Plasma Flow. In *International School and Workshop on Plasma Astrophysics*, 1981.

[15] Galvez, M. and J. Borovsky. The Expansion of Polarization Charge Layers into a Magnetized Vacuum: Theory and Computer Simulations. *Phys. Fluids B*, *8*, 1892, 1991.

[16] Gatsonis, N. and D. Hastings. A Three Dimensional Model and Initial Time Numerical Simulation for an Artificial Plasma Cloud in the Ionosphere *J. Geophys. Res.*, *96*, 7623, 1991.

[17] Goertz, C., S. Machida, and R. Smith. An Asymptotic State of the Critical Ionization Velocity Phenomenon. *J. Geophys. Res.*, *90*, 12230, 1985.

[18] Haerendel, G. Alfvén's Critical Velocity Effect Tested in Space. *Z. Naturforsch*, *37*, 728, 1982.

[19] Huba, J., H. Mitchell, J. Fedder, and P. Bernhardt. Skidding Of the Cress G-9 Barium Release. *Geophys. Res. Lett.*, *19*, 1085, 1992.

[20] Kelley M., R. Pfaff, and G. Haerendel. Electric Field Measurements During the Condor Critical Velocity Experiment. *J. Geophys. Res.*, *91*, 9939, 1986.

[21] Kim Y., K. Stephan, E. Mark, and T. Mark. Single and Double Ionization of Nitric Oxide by Electron Impact from Threshold up to 180 eV. *J. Chem. Phys.*, *74*, 6771, 1981.

[22] Lai, S., E. Murad, and W. Mcneil. An Overview of Atomic and Molecular Processes in Critical Velocity Ionization. *IEEE Trans. Plasma Sci.*, *17*, 1989.

[23] Machida, S. and C. Goertz. A Simulation Study of the Critical Ionization Velocity Process. *J. Geophys. Res.*, *91*, 11965, 1986.

[24] Machida, S. and C. Goertz. The Electromagnetic Effect on the Critical Ionization Velocity. *J. Geophys. Res.*, *95*, 11495, 1988.

[25] Machida, S., C. Goertz, and G. Lu. Simulation Study of the Ionizing Front in the Critical Ionization Velocity Phenomenon. *J. Geomagn. and Geoelectr.*, *40*, 1205, 1988.

[26] McBride, J., E. Ott, J. Boris, and J. Orens. Theory and Simulation of Turbulent Heating by the Modified Two-Stream Instability. *Phys. Fluids*, *15*, 2367, 1972.

[27] McNeil, W., S. Lai, and E. Murad. Interplay between Collective and Collisional Processes in Critical Velocity Ionization. *J. Geophys. Res.*, *95*, 10345, 1990.

[28] Mobius, E., K. Papadopoulos, and A. Piel. On the Turbulent Heating and the Threshold Condition in the Critical Ionization Velocity Interaction. *Planet. Space Sci.*, *35*, 345, 1987.

[29] Person, J., D. Resendes, H. Petschek, and D. Hastings. Effects of Collisional Processes on the Critical Velocity Hypothesis. *J. Geophys. Res.*, *95*, 4039, 1990.

[30] Squire L. and T. Baer. Cross Sections for Symmetric Charge Transfer Reactions of Xe^+ and NO^+ in Selected Internal and Translational Energy States. *J. Chem. Phys.*, *65*, 4001, 1976.

[31] Stenbaek-Nielsen, H., E. Wescott, G. Haerendel, and A. Valenzuela. Optical observations on the CRIT-II Critical Ionization Velocity Experiment. *Geophys. Res. Lett.*, *17*, 1601, 1990.

[32] Stenbaek-Nielsen, H., E. Wescott, D. Rees, A. Valenzuela, and N. Brenning. Non-solar UV Produced Ions Observed Optically from the CRIT I Critical Velocity Ionization Experiment. *J. Geophys. Res.*, *95*, 7749, 1990.

[33] Swenson, C., M. Kelley, F. Primdahl, and K. Baker. CRIT-II Electric, Magnetic, and Density Measurements within an Ionizing Neutral Stream. *Geophys. Res. Lett.*, *17*, 2337, 1990.

[34] Swenson, G., S. Mende, R. Meyerott, and R. Rairden. Charge Exchange Contamination of CRIT-II Barium CIV Experiment. *Geophys. Res. Lett.*, *18*, 401, 1991.

[35] Torbert, R., C. Kletzing, K. Liou and D. Rau. Prompt Ionization in the CRIT-II Barium Releases. *Geophys. Res. Lett.*, *19*, 973, 1992.

[36] Torbert, R. and P. Newell. A Magnetospheric Critical Velocity Experiment: Particle Results. *J. Geophys. Res.*, *91*, 9947, 1986.

[37] Wescott, E., H. Stenbaek-Nielson, T. Hallinan, H. Foppl, and A. Valenzuela. Star of Condor: A Strontium Critical Velocity Experiment, Peru, 1983. *J. Geophys. Res.*, *91*, 9933, 1986.

[38] Wescott, E., H. Stenbaek-Nielson, D. Swift, H. Foppl, A. Valenzuela and D. Rees. SR90, Strontium Shaped-Charge Critical Ionization Velocity Experiment. *J. Geophys. Res.*, *95*, 21069, 1990.

[39] Zecca A., I. Lazzizzera, M. Krauss, and C. Kuyatt. Electron Scattering from NO and N_2O below 10 eV. *J. Chem. Phys.*, *61*, 4560, 1974.

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

SMOOTHING OPERATORS FOR WAVEFORM TOMOGRAPHIC IMAGING

Prepared by: Dr. Steve Cardimona

Research Location: Phillips Laboratory,
                                      Geophysics Directorate
                                      Hanscom AFB, 01731

AF Research Colleague: Anton Dainty

Contract Number: F19628-86-C-0224

Smoothing operators for waveform tomographic imaging

Steve Cardimona* and Jan Garmany ^

26 January 1993

* Phillips Laboratory, PL/GPEH
Hanscom AFB, MA 01731
^ The University of Texas at Austin
Institute for Geophysics

ABSTRACT

The analytic kernel in the space-time domain for the Frechet derivative of
acoustic waveform data with respect to changes in the slowness model is given by
the Born approximation solution to the integral equation of waveform scattering.
Preconditioning operators in the solution of this forward problem, which may
incorporate a priori information and approximate solutions, are smoothing operators
in the imaging problem, the first iteration of a nonlinear inversion for the slowness
model. Some preconditioning operators are determined for solutions to the
parabolic wave equation, and then used to create new sensitivity functions that
retain appropriate characteristics of the true Frechet kernel in forward calculations.
The new sensitivity functions define near-source, near-receiver and far-field
kernels, as well as kernels which exhibit an amplitude decay off the ray yielding
ray-perpendicular sensitivity that scales with the Fresnel zone size. A sample
calculation from a synthetic crosswell imaging experiment shows the utility of
introducing physically appropriate model smoothing directly into the sensitivity
function of the forward problem, helping to obtain a geologically reasonable image
of the velocity model when ray coverage is insufficient.

INTRODUCTION

Tomography is used to invert seismic travel times for an estimate of the background velocities within the propagating medium (Stork and Clayton, 1991), and researchers have incorporated numerous advances into recent algorithms (Vidale, 1988, 1990; Lines and LaFehr,1989; Bregman, et al.,1989; Bortfeld, 1989; Luo and Schuster, 1991; Moser, 1991; van Trier and Symes, 1991). Methods that use the amplitude as well as the phase data from the scattered wavefield are important for increasing the resolution in imaging (Mora, 1989; Pratt and Goulty, 1991). Diffraction tomography (Devaney, 1984; Williamson, 1991) and wave-equation tomography (Woodward, 1992) move beyond traveltime inversion to focus on the information in the full waveform, and it is clear that the waveform data are sensitive to more than just the geometric ray path between the source and receiver (Cerveny and Soares, 1992). Wave equation methods are needed to model the total wavefield.

The nonlinear relationship between perturbations in the velocity model and perturbations in the waveform data can be addressed through an iterative algorithm, using the local linearization, or Frechet derivative (McGillivray and Oldenburg, 1990) of the forward scattering problem in each step. A realistic model of the earth can be inferred from an imaging experiment, the first iteration in a nonlinear inversion, given a priori knowledge of physically correct earth models in the form of model parameter constraints or covariance smoothing functions (Beydoun and Mendes, 1989). Using a preconditioning operator to help solve the integral equation of the scattering problem introduces the same operator into the linearization of the forward theory. This preconditioning operator in the forward problem incorporates exactly the data and model covariance information required for smoothing the inversion. The first section of this paper describes the relationship between the preconditioning operator in the scattering problem and the smoothing operator for imaging.

The next section of this paper shows the development of the parabolic wave equation with perturbations from an initially homogeneous slowness model. This problem is cast into its integral equation form, and the Born approximation gives

the analytic Frechet derivative for use in waveform tomographic imaging, where a source function is assumed to be known. Instead of using simply the geometric ray path (Tinti and Ugolini, 1990; Carrion, 1991; Michelena and Harris, 1991; Singh and Singh, 1991), the Frechet derivative kernel for the scattered wavefield introduces the two dimensional sensitivity function in ray-centered coordinates corresponding to a wave path (Woodward, 1992). The monochromatic Frechet kernel describes the sensitivity to perturbations in the background medium of the amplitude and the phase of the seismic signal. The shape and bandwidth of the solution after completing the forward calculation incorporates sensitivity to both transmission and wide-angle scattering within the propagating medium. .The analytic form of other sensitivity functions are introduced which include smoothing transverse to the ray while retaining appropriate physical characteristics of the true linearized solution of the forward problem. Use of these preconditioned Frechet kernels in the space-time domain imaging problem automatically applies ray-based smoothing constraints to waveform data, and this is illustrated with two imaging examples from a synthetic crosshole seismic survey. Applying appropriate image smoothing is mapped to defining preconditioning functions which incorporate a priori knowledge or approximate solutions in the forward theory, and the smoothing is accomplished in a single waveform tomographic imaging step using a modified sensitivity kernel.

MATHEMATICAL FORMALISM

Solving the integral equation of scattering

Introducing a model perturbation δm into the homogeneous differential equation defined by the differential operator L, the new wave equation for the scattering problem is $(L - \delta m(\underline{r})) \, u(\underline{r}) = 0$. The new wavefunction solution $u = u_i + u_s$ defined at position $\underline{r}$ is the sum of the incident and scattered wavefields, respectively. Since the model perturbation becomes a force term for the original wave equation, a solution of the resulting inhomogeneous equation is obtained by writing the scattered wavefield as an integral over the new force and the Green's function solution for the original equation. For a source at $\underline{r}_0$ and with the time

dependence given by t, the Green's function G_i at receiver r_j due to a secondary source (scatterer) at r' gives an integral for the scattered wavefield

$$u_s(r_j, r_0, t) = \int dr' \, G_i(r_j, t \mid r') \, \delta m(r') \, u(r', r_0, t) \, . \tag{1}$$

The result for the scattered wavefield in equation (1) is general and can be applied to a variety of surface and volume scattering problems (e.g., Kennet, 1984; Snieder, 1986). A powerful method for solving equation (1) is by iteration, i.e., putting successive approximations for u into the right hand side of the equation. This creates a Neumann series solution, and choosing $u = u_i$ in the first iteration yields the Born series. For an unperturbed equation with a delta function source, $u_i = G_i$. When the amplitude of the scattered wavefield is small relative to the incident wavefield ($u_s \ll u_i$), equation (1) can be linearized, i.e., higher order terms in u_s (in δm) are neglected. The first iteration of the Born series, or the Born approximation to u is given by

$$u_s(r_j, r_0, t) = \int dr' \, G_i(r_j, t \mid r') \, \delta m(r') \, G_i(r', t \mid r_0) = F \, \delta m \, . \tag{2}$$

The right hand side of equation (2) can be associated with the linear term in a Taylor expansion of the data as a function $u = f(m)$ of model parameter m. Equation (2) defines the Frechet derivative operator F whose kernel is given by the product of the two Green's functions. Linearizing the scattering problem implies that all multiple interactions with the scattering medium are neglected. Feynman path summation over all possible scattering paths between the source, scattering field and the receiver formalizes this (Feynman and Hibbs, 1965) and yields the same result as in equation (2) when only single scattering is considered (Clayton and Stolt, 1981). For the scattered wavefield $u_s = u - u_i = \delta u$, equation (2) can be rewritten to define the kernel D of the operator F

$$\delta u(\underset{\sim}{r}_j, \underset{\sim}{r}_o, t) = F \, \delta m = \int d\underset{\sim}{r}' \; D(\underset{\sim}{r}_j, \underset{\sim}{r}_o, t ; \underset{\sim}{r}') \, \delta m(\underset{\sim}{r}') \; . \tag{3}$$

The kernel D is the Frechet derivative operator kernel, or sensitivity function, defining the weight given to model perturbations which describes the variation in the data relative to changes in the model.

Using equation (1) for u_s, $u - u_s = u_i$ is of the general form $A(u) = u_i$ where A is a known linear operator acting on the unknown function u, and u_i is known. The first iteration of a general iterative method for solving these types of equations looks like (Kleinman and Van den Berg, 1991)

$$u \cong u_1 = u_o + \alpha_{11} C \left(u_i - A \, u_o \right) \; . \tag{4}$$

When $\alpha_{11} = 1$, and C is equal to the identity operator (I), equation (4) is the first iteration in a Neumann series; and when $u_o = u_i$ then (4) becomes the Born approximation. Kirchhoff diffraction is described completely by single scattering, thus the Born approximation gives the exact solution to this diffraction problem. The forward problem is complicated by the fact that the Born approximation is not always sufficient, and indeed the full Born series is not always convergent. In equation (4), α_{11} is a relaxation parameter for the first iteration and C is a preconditioning operator chosen to help convergence of the iterative scheme for the forward problem.

Rearranging terms in equation (4), letting $\alpha_{11} = \alpha$, and using equation (2) to define the operator A for the scattering problem of equation (1) gives a relation for $\delta u = u - u_i$

$$\delta u = \left(\alpha C F \right) \delta m \; . \tag{5}$$

The operator αC is introduced into the forward theory and modifies the true Frechet derivative operator. If αC is the identity operator, the result of equation (5) is exactly that of equation (3). As a function of source and receiver positions $\underset{\sim}{r}_o$, $\underset{\sim}{r}_j$, scattering position $\underset{\sim}{r}'$ and time t, the operator C can be designed to achieve a number

of preconditioning procedures through the operator product $(\alpha C)(F)$, including asymptotic and approximate solutions as well as data and model weighting. This corresponds to the application of a distorting function (Murch, 1992) which attempts to compensate for the difference between the incident field and the actual field.

Inverse theory and imaging

Imaging of model perturbations is the result after the first iteration in an iterative inversion algorithm. Equation (5) is in the form of a simple (nonsymmetric) linear operator equation, and hence an iteration scheme can be used to solve for the model perturbation δm as used to get equation (4) for the data. Letting the preconditioning operator for this inverse problem be the adjoint of the linear operator $(\alpha C F)$ of equation (5), and setting $\delta m_o = 0$ and $\alpha_{11} = \beta$ for a first iteration result as in equation (4), the imaging is

$$\delta m_1 = \beta \left(\alpha C F \right)^* \delta u \ . \tag{6}$$

With $\alpha=1$ and $C=I$ in the forward problem, and $\beta=1$ in the inverse problem, the imaging result of equation (6) (for F real) would look like $\delta m_1 = F^t \delta u$, and full iteration would define the normal equations for the generalized inverse. Related to linear filter theory, this imaging result is a match filter and the inversion is a deconvolutional filter for a known source wavelet when the model parameter is the reflectivity of the medium (Cardimona, 1991). Related to single-step migration/inversion (Beydoun and Mendes, 1989), the image is a migration of the data, and the inversion result is given by the deconvolution of the image by a Hessian correction.

With a nontrivial C operator, the imaging solution given by equation (6) becomes a modified gradient solution (Tarantola, 1984) where the preconditioning operator helps to accelerate convergence and/or simplify the computations. The result in equation (6) relates exactly to that obtained after the first iteration of a conjugate gradient algorithm. In order to minimize a weighted combination of

prediction error and model covariance size, the first iteration of a gradient method for determining changes in model parameter m is proportional to (Tarantola, 1984; Carrion, 1989)

$$\delta m_1 = C_m F^* C_u^{-1} \delta u \, ,$$

(7)

where F^* is the adjoint of F ($F^* = F^t$ for real F), and C_u and C_m are the least squares functional weighting operators, equivalent to the data covariance and model covariance matrices for the discrete problem (Tarantola, 1986). The right hand side of equation (7) defines the adjoint operator (Tarantola, 1987) acting on perturbations in the data as used in equation (6), and shows that the preconditioning operator C of the forward problem can be defined to contain within it the data weighting given by C_u and the model space smoothing given by C_m. Introducing C as a weighting function with respect to source and receiver positions, scatterer position, frequency or time can help to suppress the influence of noisy data on the inversion, alter the expected resolution for the model parameter or focus the inversion on a specific range of frequency, offset or time.

With no preconditioning operator, a model perturbation estimate is of the form $\delta m_1 = \delta m_{est} = F^t \delta u$. The kernel for the transpose operator F^t in the inverse problem is the same as that of F in the forward problem, but the operation is over the transposed variables. Using equation (3) with a receiver r_j, source at $\underset{\sim}{r}_n$ and time dependence given by t, the model perturbation estimate at $\underset{\sim}{r}'$ is

$$\delta m_{est}(\underset{\sim}{r}') = \sum_{n,j} \int_0^T dt \ D(\underset{\sim}{r}_j, \underset{\sim}{r}_n, t ; \underset{\sim}{r}') \ \delta u(\underset{\sim}{r}_j, \underset{\sim}{r}_n, t) \ .$$

(8)

The transpose operator in equation (8) incorporates a contraction over the time variable within the data window T, and this correlation operation becomes the imaging condition. The summation (stack) over all source and receiver positions incorporates the information from all the data traces available. A nontrivial preconditioning operator C in equation (5) defines a new sensitivity function for the

forward problem and thus introduces a smoothing function into the imaging equation (8) and the inverse problem.

PARABOLIC WAVE EQUATION APPLICATION

Forward scattering problem

Starting from the scalar Helmholtz equation with wavenumber k, $\phi_{xx} + \phi_{yy} + k^2 = 0$, let $\phi(x,y) = \exp(i\,k_o x)\,P(x,y)$ define the wavenumber k_o associated with the initially homogeneous background medium. For frequency ω, introduce slowness s as the model parameter such that $\omega s = k$. With a background slowness s_o, let $(s-s_o) = \delta s$ be the perturbation in the model. Putting ϕ into the Helmholtz equation gives a differential equation for $P(x,y)$ in which the P_{xx} term is neglected. Noting that $(k^2 - k_o^2) = \omega^2(2s_o\delta s + \delta s^2)$ and neglecting the δs^2 term, the parabolic equation for $P(x,y)$ is given by

$$P_{yy} + i\,2\omega\,s_o\,P_x = -2\omega^2\,s_o\,\delta s\,P \ . \tag{9}$$

Applying the parabolic wave equation requires the use of relatively high frequencies and rays that diverge relatively slowly (Graves and Clayton, 1990; Wapenaar, 1990). When $\delta s = 0$ (free particle), introduction of a delta function source gives the exact Green's function solution (Mathews and Walker, 1970) needed to solve the inhomogeneous equation (9). After linearizing the solution for P as in equations (2) and (3) (with frequency ω)

$$P - P_o = P_1 = \int dx' \int dy' \ i\,\omega\,P_o\,K(x,y,\omega\,;\,x',y')\ \delta s(x',y') \ , \tag{10}$$

where $[i\,\omega P_o K'(\omega,x',y')]$ is the monochromatic Frechet derivative operator kernel for this problem, acting on slowness perturbations $\delta s(x',y')$ to give perturbations $P_1(x,y,\omega)$ in the data. Note that equation (10) is set up directly for extension to the Rytov approximation (Devaney, 1984; Rajan and Frisk, 1989), where the scattering is associated with a complex phase term $\psi_s = P_1/P_o$. Since the

Rytov approximation can be achieved through an extension of the Born theory, this paper is restricted to the study of the function K of equation (10) which defines the weight applied to the model perturbation at point (x',y'), as felt at the receiving point (x,y).

Without loss in generality, the rectangular coordinate system is translated so that the source is at (0,0), and rotated so that the x-axis is the linear ray path between the source and receiver, putting the receiver at the point (X,0). After making this change of variables, the function K for the Born approximation solution in (10) is in the form

$$K(X,\omega \; ; x',y') = \frac{1}{H} \sqrt{\frac{1}{2\pi}} \; \exp \left\{ \; i \; \left[\frac{(y')^2}{2H^2} - \frac{\pi}{4} \right] \; \right\} \; . \tag{11}$$

H has the units of distance and controls the amplitude variation in the ray-parallel direction, giving the correct singularities at the source and receiver

$$H = \sqrt{\frac{(X-x') \; x'}{X \, \omega \, s_0}} \; . \tag{12}$$

When $\delta s = 0$ in equation (10) the scattered wavefield $P_1 = 0$, and $P = P_0$ the free space solution. When the perturbation in the background slowness is a constant $\delta s = \delta s_0$, the integral of (10) becomes, after making the change of variables described above and using equations (11) and (12),

$$\frac{P_1}{P_0} = i \, \omega \sqrt{\frac{-i}{2\pi}} \; \delta s_0 \int_0^X \frac{dx'}{H} \int_{-\infty}^{\infty} dy' \; \exp\left[\frac{i}{2H^2} (y')^2 \right] \; . \tag{13}$$

The integral of (13) is solved with the help of a useful definite integral

$$\int_{-\infty}^{\infty} dx \ e^{-ax^2} = \sqrt{\frac{\pi}{a}} \ .$$ (14)

With (14), equation (13) becomes

$$\frac{P_1}{P_o} = i\,\omega\,\delta s_o \sqrt{\frac{-i}{2\pi}} \int_0^X \frac{dx'}{H} \sqrt{\frac{-2H^2\pi}{i}} = i\,\omega\,\delta s_o \int_0^X dx' = i\,\omega\,\delta s_o\, X \ .$$ (15)

Equation (15) gives the correct phase change due to the perturbed slowness field, as seen from

$$\left(\frac{P}{P_o} - 1 \right) = \left\{ \exp\left[i\,(k-k_o)X \right] - 1 \right\} = i\,\omega\,\delta s_o\, X + O(|\delta s_o|^2) \ .$$ (16)

With a linear trend in y', there is no perturbation in the phase or amplitude of the data P, since the y'-integral of the odd linear function against the even Gaussian y' dependence in the function K is zero. This is correct for X small compared to the radius of curvature of the perturbed ray. For a quadratic in y' as a slowness perturbation given by $\delta s(x',y') = \alpha\,(y')^2/2$, then equation (10) becomes

$$\frac{P_1}{P_o} = i\,\omega \sqrt{\frac{-i}{2\pi}}\, \frac{\alpha}{2} \int_0^X \frac{dx'}{H} \int_{-\infty}^{\infty} dy' \exp\left[\frac{i}{2H^2}(y')^2 \right] (y')^2 \ .$$ (17)

The integral in (17) is solved using another useful definite integral which can be obtained from (14)

$$-\frac{\partial}{\partial a}\left(\int_{-\infty}^{\infty} dx\, e^{-ax^2} \right) = \int_{-\infty}^{\infty} dx\, e^{-ax^2} x^2 = \frac{\sqrt{\pi}}{2}\, a^{-3/2} \ .$$ (18)

Using (18), equation (17) becomes

63

$$\frac{P_1}{P_o} = \frac{i\,\omega\alpha}{4}\sqrt{\frac{-i}{2}}\int_0^X \frac{dx'}{H}\left(-\frac{2}{i}\right)^{3/2} H^3 = -\frac{\alpha}{2Xs_o}\int_0^X dx'\left[Xx' - (x')^2\right]\ ,\qquad (19)$$

so that

$$\frac{P}{P_o} = 1 + \frac{P_1}{P_o} = 1 - \frac{\alpha\,X^2}{12\,s_o}\ .\qquad\qquad (20)$$

When the second derivative of the slowness is negative ($\alpha<0$, as in a low velocity acoustic channel), equation (20) gives the correct amplitude focusing described by the geometrical spreading equations of ray theory (Cerveny and Hron, 1980). Figure 1 shows the amplitude and wrapped phase for the function of equation (11). From the phase variation in the transverse coordinate, it is clear the ray feels a region around the classical ray path (Woodward, 1992), and the scale of this region relates to the Fresnel zone size. The amplitude of this kernel is more curious, having no variation in the transverse coordinate at all.

In the following discussion, some integral operators are defined which, when applied to the exact Frechet derivative kernel given by equation (11), create new sensitivity kernels for the forward problem that retain the appropriate analytic scattering results of equations (13)-(20). With a transformation operator kernel given by σ, a new function K_s is defined by

$$K_s(X,\omega;x',y') = \int_0^X dx'' \int_{-\infty}^{\infty} dy''\ K(X,\omega\ ;\ x'',y'')\ \sigma(x'',y'',\omega\ ;\ x',y')\ ,\qquad (21)$$

so that a new form of equation (10) is

$$\frac{P_1}{P_o} = \int_0^X dx' \int_{-\infty}^{\infty} dy'\ i\,\omega\ K_s(X,\omega\ ;\ x',y')\ \delta s(x',y')\ .\qquad\qquad (22)$$

If $\sigma(x'',y'',\omega;x',y') = \delta(x''-x')\delta(y''-y')$, then $K_s=K$ and (22) becomes exactly (10). Note that σ could be a function of source and receiver positions as well, and then (21) would define an operator over the (x,y) coordinates of (10). Using equations (21) - (22) and interchanging the order of integration shows that the function σ acting on the Frechet kernel is actually a smoothing function on the model parameter changes $\delta s(x',y')$ (Cardimona, 1993).

With the results of equations (15) and (20) as criterion for defining σ in (21) and (22), the analytic expression of a small class of smoothing functions can be determined and is given by

$$\sigma(x'',y'',\omega;x',y') = \delta(x''-x') \sqrt{\frac{1}{2\pi} \frac{1}{F}} \exp\left(\frac{y'}{FH} y'' - i\frac{\pi}{4}\right), \qquad (23)$$

where H is the same as in equation (12), and F is another function of X, x', s_0 and ω. The smoothing operator of equation (23), when used as in equation (21) to get a new sensitivity kernel, exchanges H with the function F in equation (11). That is, with (23), K_s is the same as K in equation (11) except that the function F describes the ray-parallel amplitude dependence. When $F = H$ in equation (23), then $K_s = K$. Introducing three other possible functions for F that still retain the correct ray theory scattering results in the discussion of equations (13)-(20) defines different ray-parallel dependence

$$F = \sqrt{\frac{x'}{3\omega s_0}}, \quad F = \sqrt{\frac{(X-x')}{3\omega s_0}} \quad \text{and} \quad F = \sqrt{\frac{X}{6\omega s_0}}, \qquad (24)$$

describing near-source, near-receiver, or far-field x'-dependence in the sensitivity kernel, respectively.

Another class of smoothing functions, again determined using equations (15), and (20)-(22), give interesting sensitivity kernels. With H and F given by (12) and (24),

$$\sigma(x'',y'',\omega;x',y') = \delta(x''-x') \sqrt{\frac{1}{2}} \frac{1}{F} \exp\left[-\left(\frac{H}{F} \frac{y'}{y''} \right)^2 - i\frac{\pi}{4} \right] .$$ (25)

Using the helpful definite integral

$$\int\limits_{-\infty}^{\infty} dx \, \exp\left(-ax^2 - \frac{b}{x^2}\right) = \sqrt{\frac{\pi}{a}} \, \exp(-2\sqrt{ab}) ,$$ (26)

equations (11) and (25) together with equation (21) yields

$$K_s(X,\omega;x',y') = \frac{1}{F}\sqrt{\frac{1}{2}} \exp\left(-\frac{|y'|}{F} \right) \exp\left[i \left(\frac{|y'|}{F} - \frac{\pi}{4} \right) \right] .$$ (27)

With F=H in equation (25), equation (27) is a sensitivity function with the correct ray-parallel amplitude dependence, but including an exponential damping term perpendicular to the ray. The physics of wave propagation details that the phase contributions to the integration in equations (10) and (22) from the y' coordinate (ray-perpendicular) damp out away from the ray due to incoherent stacking away from the stationary path. The amplitude of the kernel given by (11) does not reflect the Fresnel zone sampling of the ray. Incorporating the amplitude decline away from the classical ray, the kernel of (27) retains sensitivity relative to the Fresnel zone width while putting limits on the ray-perpendicular integration for the terms in the Born series (Figure 2).

The functions given by equations (23) and (25) are preconditioning operators in the forward problem to help the Born approximation be more accurate, or help the Born series converge (Cardimona, 1993). For example, the far-field kernel has no square-root singularities at the source and receiver, thus allowing the forward iterative method to be more stable. In the imaging problem, the preconditioning functions work to smooth out the sensitivity variation of the Born kernel.

Inverse problem

For the imaging problem, a solution in the form of equation (8) is sought. Referring back to the original inhomogeneous equation (9), let

$$\phi_1 = e^{i k_o X_j} P_1(X_j, 0) \equiv \delta u(\underset{\sim}{r}_j, t) \tag{28}$$

be the perturbation in the time-dependent data for a source and receiver pair defined by position $\underset{\sim}{r}_j$ and distance X_j. This particular forward problem is then given by

$$\delta u(\underset{\sim}{r}_j, t) = \int dx' \int dy' \; \kappa(\underset{\sim}{r}_j, t \; ; \; x', y') \; \delta s(x', y') \; . \tag{29}$$

In equation (29) the forward response for a set of frequencies is calculated, and then the Fourier transform is applied to get the time dependent kernel κ

$$\kappa(\underset{\sim}{r}_j, t \; ; \; x', y') = \int d\omega \, e^{i \omega t} \left[i \, \omega \, W(\omega) \, e^{i k_o X_j} P_o(X_j, 0) \; K(X_j, \omega \; ; \; x', y') \right] \; . \tag{30}$$

$W(\omega)$ is the source wavelet signature, and K is a sensitivity function from the linearization of the forward problem. The transpose of the kernel κ in equation (29) corresponds to the application of the same sensitivity function on the data, integrating over the transposed variables. Comparing with equation (8), the transpose operation involves an integral over time and a summation over all source and receiver combinations

$$\delta s(x', y') = \sum_j \int_0^T dt \; \kappa(\underset{\sim}{r}_j, t \; ; \; x', y') \; \delta u(\underset{\sim}{r}_j, t) \; . \tag{31}$$

67

RESULTS

Synthetic crosswell imaging

Figure 3a shows the geometry for a synthetic scatterer imaging experiment in a simple crosshole waveform tomography problem. Five sources at 50 m spacing in well #1 and five receivers at 50 m spacing in well #2 gives 25 source-receiver pair records. For δs a point perturbation in the middle of the model at (.25,.10), the calculated gradients of the model update after a single iteration are shown in Figures 3 and 4, the results using the exact sensitivity kernel K from equation (11) and a smoothed version K_s from equation (27), respectively. The synthetic seismograms are calculated for perturbation from a constant velocity model in a 10-60 Hz band, for each of the five shots and five receivers using equations (10)-(12) and (28)-(30). Then, using equation (31) for each of the sensitivity functions yields images of the slowness perturbation.

Comparing Figures 3d and 4c, the Born_gradient result is clearly the best; although it still shows some horizontal smearing of the point perturbation due to lack of surface data. The inversion damping associated with the use of K_s is most pronounced where data coverage is sparse, as seen in the single shot gather imaging of Figure 4a&b and near each well in the full gradient results of Figure 4c, compared with Figures 3c-d, respectively. It is clear from the selected shot gathers (Figure 3b&c and 4a&b) that using the smoothed sensitivity function K_s does the job of damping the inversion, yielding qualitatively more desirable results. Figure 5 illustrates this through another example of the reconstruction of a constant background slowness perturbation (Figure 5a) using only five records from a shot at position #3 at depth 100 m. (as in Figures 3c and 4b). The gradient calculations have been modified by a function which helps to make up for the known poor ray coverage (Figure 5b). The imaging result using K_s (Figure 5d) is clearly the better reconstruction of the constant perturbation, in contrast to the image using the true Born kernel (Figure 5c).

DISCUSSION

In imaging experiments, the variation in model parameters is sought, and geologically reasonable results are achieved with the help of smoothing functions.

Preconditioning operators in the forward theory incorporating approximate solutions and a priori information become smoothing operators in the imaging problem. Using analytically determined sensitivity functions in ray centered coordinates for perturbations to the parabolic equation, this paper develops a specific application to imaging using full waveform data in a tomographic experiment where only phase information (traveltime data) is often used. The analytic extension of these ideas to more iterations in the inversion is not trivial. In waveform inversion, it is the action of the adjoint operator to the forward problem acting on the data residuals which yields the required update in the model. When the adjoint is other than the exact transpose of the forward (Born approximation) operator, it acts to smooth the gradient in the inverse calculation in order to help the convergence of the iterative scheme. It is from the solution to the forward problem that physically reasonable smoothing operators for the inverse problem can be determined. The smoothing operators developed in this paper were defined for specific cases to correspond with the Born approximation results in the forward calculations. However, any type of smoothing normally helps convergence, and physically reasonable operators defined from approximate or asymptotic solutions in the forward theory are enticing and easily introduced directly into the imaging step.

ACKNOWLEDGEMENTS

REFERENCES

Beydoun, W. B., and Mendes, M., 1989, Elastic ray-Born l_2-migration/inversion, Geophysical Journal, 97, 151-160.

Bortfeld, R., 1989, Geometrical ray theory, rays and traveltimes in seismic systems (second-order approximations of the traveltimes): Geophysics, 54, 342-349.

Bregman, N. D., Bailey, R. C., and Chapman, C. H., 1989, Crosshole seismic tomography: Geophysics, 54, 200-215.

Cardimona, S., 1991, Waveform inversion and digital filter theory: Geophysics, 56, 534-536.

Cardimona, S., 1993, Integral equations and scatterer imaging, submitted to Inverse Problems.

Carrion, P. 1989, Generalized non-linear elastic inversion with constraints in model and data spaces: Geophysical Journal, 96, 151-162.

Carrion, P., 1991, Dual tomography for imaging complex structures: Geophysics, 56, 1395-1404.

Cerveny, V., and Hron, F., 1980, The ray series method and dynamic ray tracing system for three-dimensional inhomogeneous media: Bull. seism. Soc. Am., 70, 47-77.

Cerveny, V., and Soares,, J. E. P., 1992, Fresnel volume ray tracing: Geophysics, 57, 902-915.

Clayton, R. W., and Stolt, R. H., 1981, A Born-WKBJ inversion method for acoustic reflection data: Geophysics, 46, 1559-1567.

Devaney, A. J. 1984, Geophysical diffraction tomography: IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-22, No.1, 3-13.

Feynman, R. P., and Hibbs, A. R. 1965, Quantum mechanics and path integrals: McGraw-Hill, Inc.

Graves, R. W., and Clayton, R. W., 1990, Modeling acoustic waves with paraxial extrapolators: Geophysics, 55, 306-319.

Kennett, B. L. N., 1984, An operator approach to forward modelling, data processing and migration: Geophys. Prosp., 32, 1074-1090.

Kleinman, R.E., and Van den Berg, P.M., 1991, Iterative methods for solving integral equations: Radio Science, 26, 175-181.

Lines, L. R., and LaFehr, E. D., 1989, Tomographic modeling of a cross-borehole data set: Geophysics, 54, 1249-1257.

Luo, Y., and Schuster, G. T., 1991, Wave-equation traveltime inversion: Geophysics, 56, 645-653.

McGillivray, P. R., and Oldenburg, D. W., 1990, Methods for calculating Frechet derivatives and sensitivities for the non-linear inverse problem, a comparative study: Geophys. Prosp., 38, 499-524.

Mathews, J., and Walker, R. L., 1970, Mathematical methods of physics: Benjamin/Cummings Publishing Co.

Michelena, R. J., and Harris, J. M., 1991, Tomographic traveltime inversion using natural pixels: Geophysics, 56, 635-644.

Mora, P., 1989, Inversion = migration + tomography: Geophysics, 54, 1575-1586.

Moser, T. J., 1991, Shortest path calculation of seismic rays: Geophysics, 56, 59-67.

Murch, R.D., 1992, An extended Born approximation: Inverse Problems, 8, L5-L11.

Pratt, R. G., and Goulty, N. R., 1991, Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data: Geophysics, 56, 208-224.

Rajan, S. D., and Frisk, G. V. 1989, A Comparison between the Born and Rytov approximations for the inverse backscattering problem: Geophysics, 54, 864-871.

Singh, R. P., and Singh, Y. P., 1991, RAYPT - A new inversion technique for geotomographic data: Geophysics, 56, 1215-1227.

Snieder, R., 1986, 3-D linearized scattering of surface waves and a formalism for surface wave holography: Geophys. J. R. astr. Soc., 84, 581-605.

Stork, C., and Clayton, R. W., 1991, Linear aspects of tomography velocity analysis: Geophysics, 56, 483-495.

Tarantola, A., 1984, Linearized inversion of seismic reflection data: Geophys.Prosp., 32, 998-1015.

Tarantola, A., 1986, A strategy for non-linear elastic inversion of seismic reflection data: Geophysics, 51, 1893-1903.

Tarantola, A., 1987, Inverse problem theory, methods for data fitting and model parameter estimation: Elsevier Science Publishing Company, Inc.

Tinti, S., and Ugolini, S., 1990, Pre-selection of seismic rays as a possible method to improve the inverse problem: Geophys. J. Int, 102, 45-61.

van Trier, J., and Symes, W. W., 1991, Upwind finite-difference calculation of travel times: Geophysics, 56, 812-821.

Vidale, J., 1988, Finite difference calculation of traveltimes: Bull. Seis. Soc. Am., 78, no. 6, 2062-2076.

Vidale, J., 1990, Finite-difference calculation of traveltimes in three dimensions: Geophysics, 55, 521-526.

Wapenaar, C. P. A., 1990, Representations of seismic sources in the one-way wave equations: Geophysics, 55, 786-790.

Williamson, P. R., 1991, A guide to the limits of resolution imposed by scattering in ray tomography: Geophysics, 56, 202-207.

Woodward, M. J., 1992, Wave-equation tomography: Geophysics, 57, 15-26.

FIGURE CAPTIONS

Fig. 1. The amplitude and (wrapped) phase in ray-centered coordinates for the monochromatic Frechet derivative of the parabolic wave equation: (a) phase at 10 Hz; (c) phase at 60 Hz. The amplitude variation in (b) is qualitatively the same for both frequencies, differing by a scalar. (d) The geometry for a source at (0,0) and a receiver at distance $X=0.5$ km from the source. All spatial coordinates in km.

Fig. 2. Plots of the amplitude and (wrapped) phase for a preconditioned sensitivity function associated with the parabolic wave equation at specific frequencies: (a) and (b) 10 Hz; (c) and (d) 60 Hz. The ray-centered coordinate geometry is as in Figure 1d.

Fig. 3. (a) Geometry of a simple cross borehole tomographic experiment. (b)-(d) Example gradient calculation (imaging experiment) using the true Born kernel for a point perturbation in the model. (b) and (c) are the contributions to the gradient from the record gathers for shot #1 and for shot #3, respectively. The full gradient update (the image) of perturbations to the background slowness is given by a sum of the contributions from all 25 data records in (d).

Fig. 4. Example gradient calculation (imaging experiment) using the smoothed sensitivity function K_s for a point perturbation in the model. The cross-hole geometry is as in Figure 3a, with horizontal and vertical scales in km.

Fig. 5. (a) The slowness perturbation to the constant initial model for the forward synthetics is taken as a small constant positive value. The cross-hole geometry is as in Figure 3a. (b) The ray coverage through the medium for the single shot gather, used to modify the gradient calculations. The darker the plot, the less ray coverage. (c) Image using the Born kernel and modified by a function as in (b) describing the poor ray coverage. (d) Modified image using the smoothed kernel K_s.
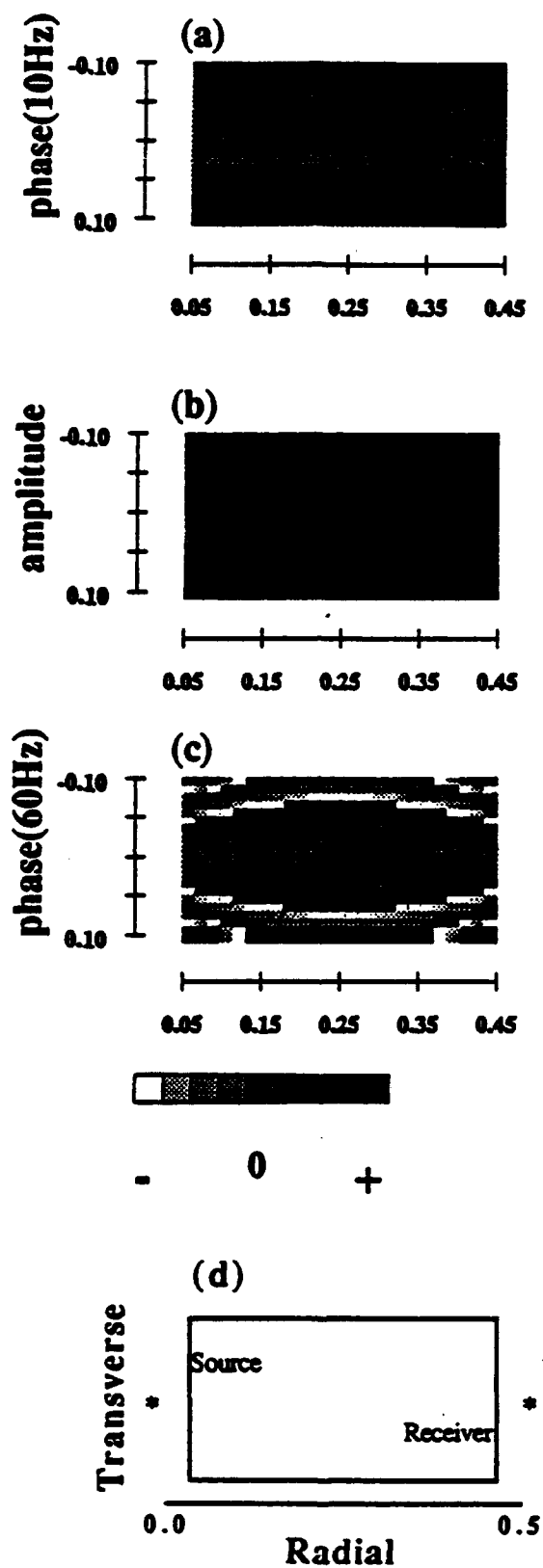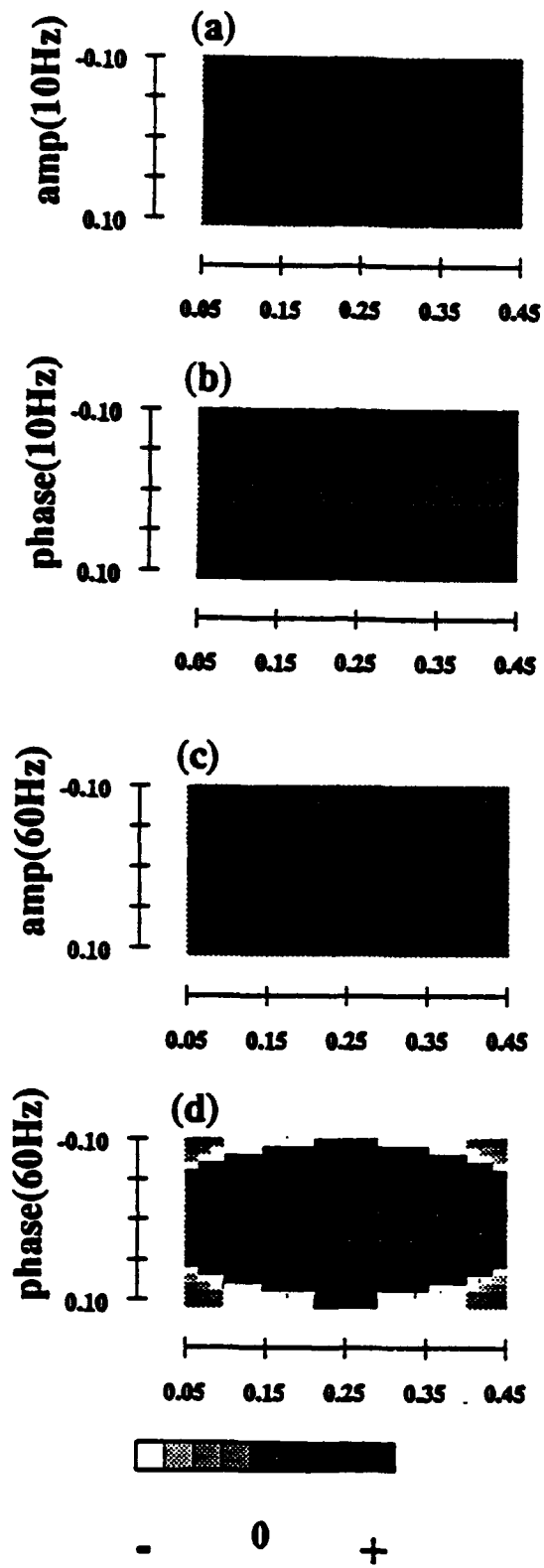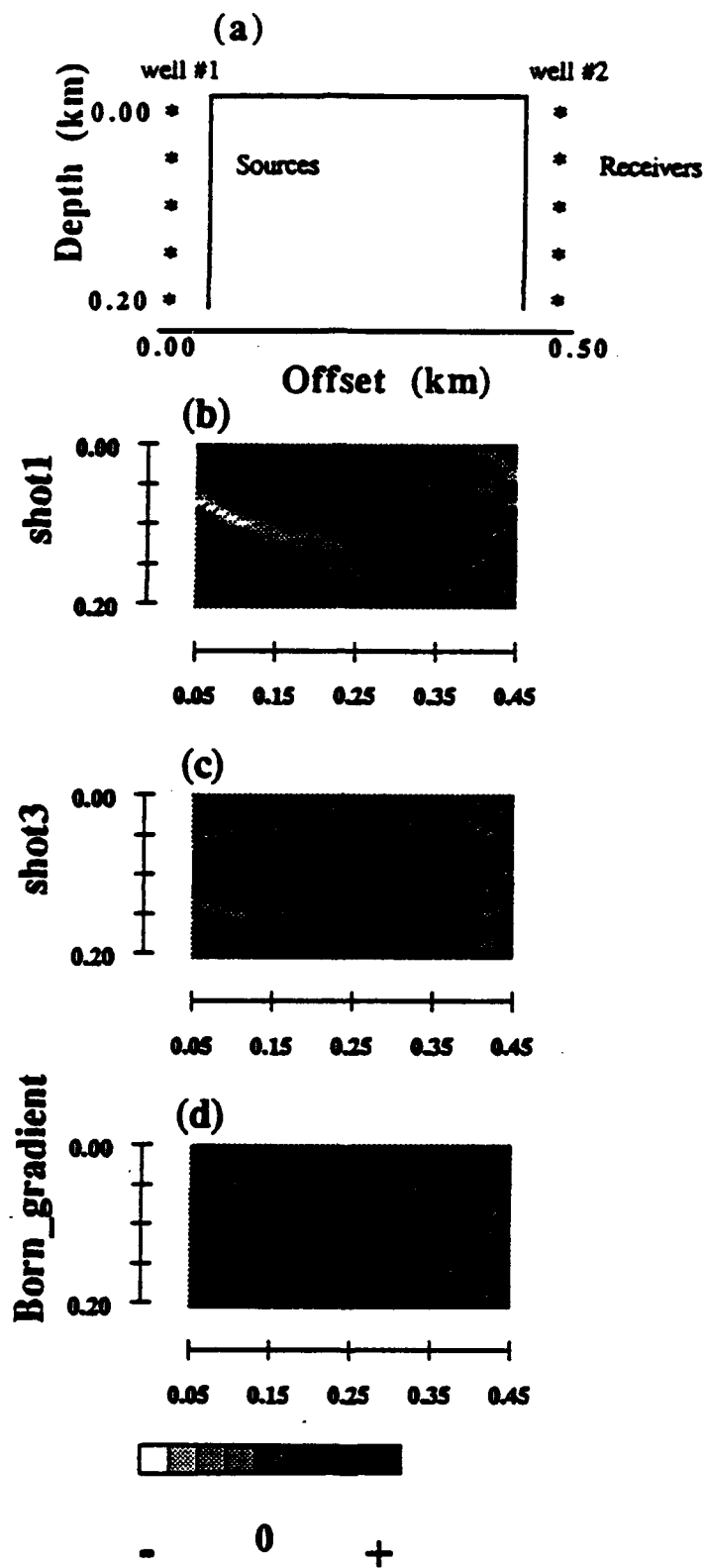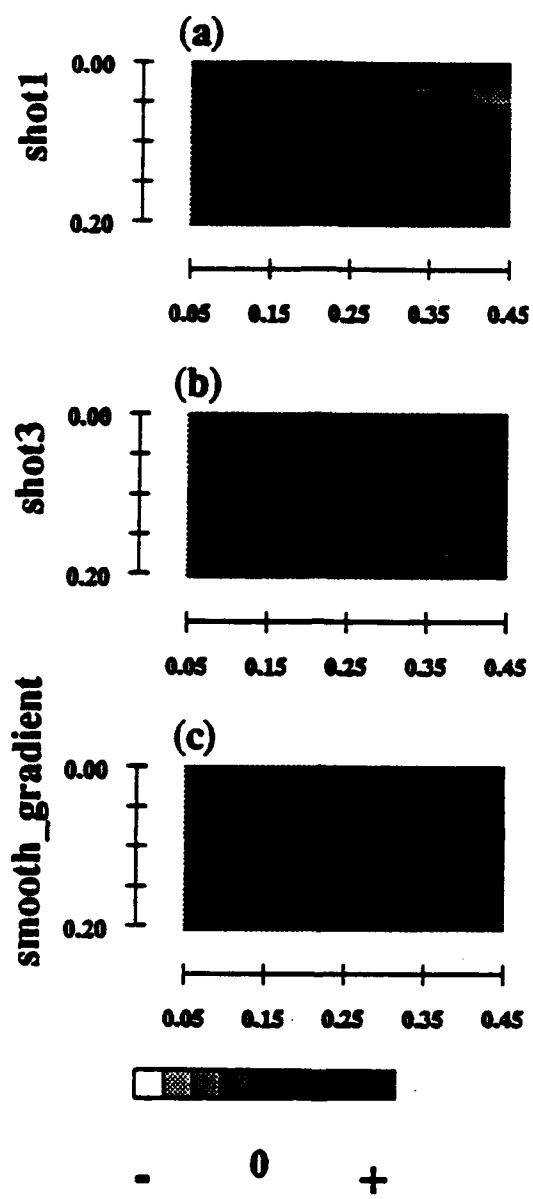
FIGURE 1

FIGURE 2

(a)

well #1   well #2

Depth (km)

0.00 *   *

Sources   * Receivers

0.20 *   *

0.00   0.50

Offset (km)

(b)

shot1

0.00

0.20

0.05   0.15   0.25   0.35   0.45

(c)

shot3

0.00

0.20

0.05   0.15   0.25   0.35   0.45

(d)

Born_gradient

0.00

0.20

0.05   0.15   0.25   0.35   0.45

-          0   +

FIGURE 3

(a) shot1

(b) shot3

(c) smooth_gradient

FIGURE 4

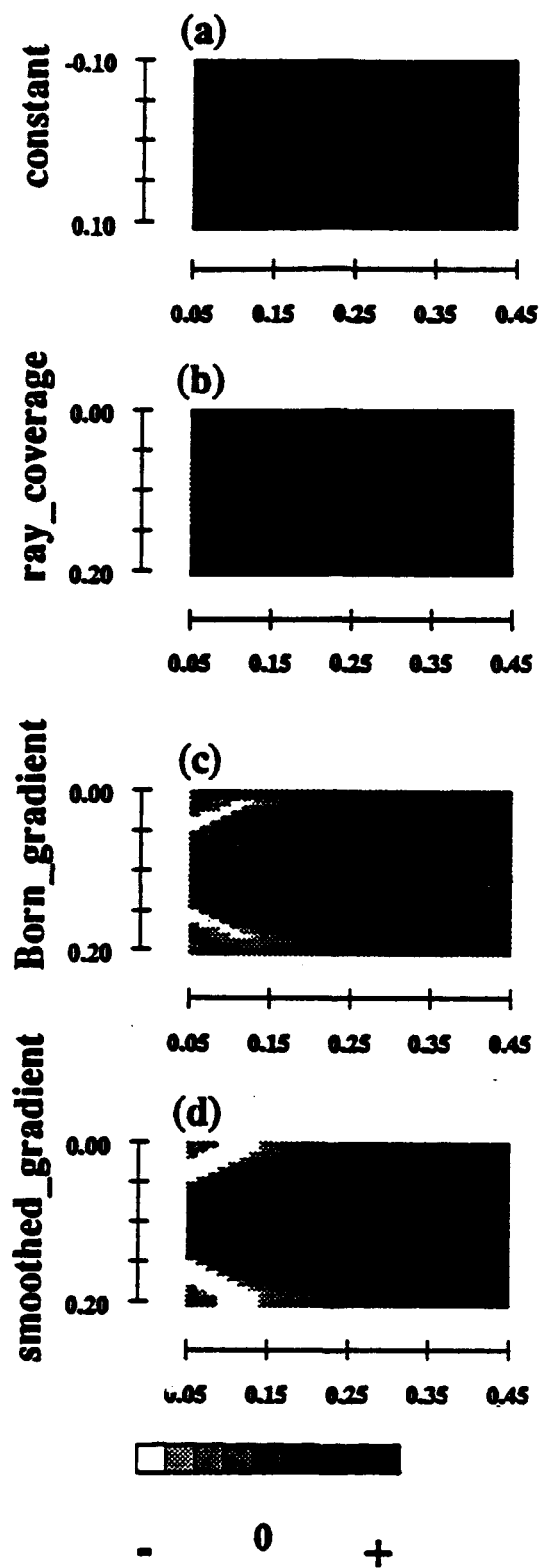FIGURE 5

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

INTEGRAL EQUATIONS AND SCATTERER IMAGING

Prepared by: Dr. Steve Cardimona

Research Location: Phillips Laboratory,
                                      Geophysics Directorate
                                      Hanscom AFG, MA 01731

AF Research Colleague: Anton Dainty

Contract Number: F19628-86-C-0224

Integral equations and scatterer imaging

Steve Cardimona

Phillips Laboratory/GPEH

29 Randolph Road

Hanscom AFB, MA 01731-3010

E-mail: cardi@doc.plh.af.mil

Office phone: (617) 377 - 2652

29 January 1993

Abstract

The preconditioning operator used in solving the integral equation of the forward scattering problem is shown to be a smoothing operator introduced into a calculus-based optimization technique to create an image of model perturbations. Some preconditioning operators for solutions to the paraxial wave equation are analytically determined from the physics of the forward scattering. The qualitative nature of these preconditioning operators is described showing how they become smoothing functions on the model space of slowness perturbations.

Integral equations that arise in scattering problems describe the wavefunction (data) in terms of a perturbation in the potential. The Neumann series is one example of more general iterative methods which can be employed to get a solution to the forward problem. The particular iteration method by which the forward problem is solved has direct significance when looking at the linearization of that solution for use in the inverse problem. With the operators defined in physical space, the intuitive description of the preconditioning operator used in solving the forward scattering problem defines appropriate smoothing operators in scatterer imaging, the first iteration result in an optimization technique to solve the nonlinear inverse problem.

Introducing a model perturbation δm into a differential equation defined by the operator D yields the forward scattering problem for wavefunction ϕ at positions r

$$(D - \delta m(r)) \, \phi(r) = 0 , \tag{1}$$

where the solution $\phi = \phi_i + \phi_s$ is the sum of the incident (background) and scattered wavefields. Equation (1) can be solved by writing the scattered wavefield as an integral over the Green's function solution to the original, unperturbed wave equation with a delta function source. Letting this Green's function be G_o, a solution can be written as, for a receiver at r and scattering points r' in the medium,

$$\phi(r) = \phi_i(r) + \int dr' \, G_o(r|r') \, \delta m(r') \, \phi(r') \equiv \phi_i + K \phi . \tag{2}$$

The scattered wave field in (2) is defined by the integral operator K. Designating the operator (I-K) to be L (I is the identity operator) gives an integral equation of the second kind

$$L \, \phi = \phi_i . \tag{3}$$

A general iterative scheme [1] for solving (3) is given by

$$\phi_n = \phi_{n-1} + \sum_{m=1}^{n} \alpha_{nm} C \left(\phi_i - L \phi_{m-1} \right) .$$ (4)

In (4), C is a preconditioning operator, and the α_{nm} are coefficients to help convergence of ϕ_n to the solution ϕ. Equation (4) yields the Born series solution when C=I, the identity operator; α_{nm}=0, m<n; α_{nn}=1; and $\phi_0 = \phi_i$. Letting $\alpha_{11}=\alpha$, the first iteration of (4) is

$$\phi \cong \left(I + \alpha C K \right) \phi_i .$$ (5)

With α=1 and C=I, (5) is the Born approximation which describes single scattering within the medium. If the Born series is convergent, it corresponds to the geometric series expansion of the formal solution of (2) and describes the resolvent R from the Fredholm solution to the integral equation, $\phi=(I+R) \phi_i$ [2]. In the Born approximation, the true Green's function of the resolvent R is approximated by the Green's function from the original homogeneous equation given by the operator K in (2). For example, Kirchhoff diffraction is described completely by single scattering, and the Born approximation gives the exact solution to this problem.

Complicating the forward problem is fact that the Born approximation is not always sufficient, and the full Born series is not always convergent. For the Born power series in K obtained from (4), a comparison series to ϕ_n can be introduced,

$$|\phi_n| \leq |\phi_i|_{max} \, |G_o|_{max}^n \, |\delta m|_{max}^n \, |r_2 - r_1|^n ,$$ (6)

where $|r_2 - r_1|$ is symbolic for the size of the domain of the kernel $(G_o \, \delta m)$ of K. Then, by the ratio test, convergence of the Born series is assured if

$$|G_o|_{max} \; |\delta m|_{max} \; |r_2 - r_1| \; < \; 1 \; . \tag{7}$$

For a norm $\| \cdot \|$ on the Hilbert space where K of (2) is defined, the condition (7) is related to the convergence condition given by the spectral radius $\sigma(K)$

$$\sigma(K) = \lim_{n \to \infty} \| K^n \|^{\frac{1}{n}} \; < \; 1 \; . \tag{8}$$

When the Born series is not convergent, a preconditioning operator introduced via (4) can help to achieve convergence [1]. The operator C in (5) corresponds to the distorting function [3] which attempts to compensate for the difference between the incident field and the actual field. This preconditioning operator can help the linearization to better approximate the resolvent, thus helping convergence where the constant α has only limited effect. A relation similar in form to (7) can qualitatively describe a convergence criterion after the first approximation in (5),

$$|G_o'|_{max} \; |\delta m|_{max} \; |r_2 - r_1| \; < \; 1/\alpha \; . \tag{9}$$

Here $(G_o' \; \delta m)$ is the kernel of the operator $[C \; K]$ of (5). The parameter α relaxes the size restriction on the magnitude of δm, and the operator C is designed to restrict the magnitude range of the Green's function and/or the size of its domain of integration.

The calculus-based optimization problem for determining the perturbation in the model δm from changes in the recorded data δd relies on an equation such as (5)

$$\delta d = \phi - \phi_i \cong \left(\alpha \, C \, K \right) \phi_i \cong \left(\alpha \, C \, F \right) \delta m \; . \tag{10}$$

F is a new integral operator defined by the product of the Green's function of the original homogeneous equation and the incident wavefield. The right hand side

of (10) is associated with the linear term in a Taylor expansion of the data with respect to changes in the model. Letting $\alpha = 1$ and $C = I$, the operator F is exactly the Frechet derivative.

Noting that (10) is similar in form to (3), a solution can be obtained for δm with an iteration scheme as in (4). Letting the preconditioning operator for this inverse problem be the adjoint of the linear operator [α C F] from the forward theory (since the inverse problem is not symmetric), and setting $\delta m_o = 0$ and $\alpha_{11} = \beta$, the imaging result δm_1 is

$$\delta m_1 = \beta \left(\alpha C F \right)^* \delta d . \tag{11}$$

With $\alpha = 1$ and $C = I$ in the forward problem, $\beta = 1$ in the inverse problem and F real, the imaging result of equation (11) is $\delta m_1 = F^* \delta d$. For this special case, the imaging result corresponds to a match filter, and the full inversion becomes a linear deconvolution of a known source wavelet, when the model parameter is the reflectivity of the medium [4]. Assuming the existence of α for convergence of this linear inversion, the appropriate scaling factor is equivalent to normalizing by the zero lag term of the source autocorrelation and by the time window length of the data trace which is the domain of F in this example.

In obtaining analytic results, the operator C of (4) is often taken to be the identity for a symmetric problem and the adjoint of L for the nonsymmetric case [1]. For computational purposes, preconditioning is most often determined through matrix methods [5]. The preconditioning operator C can be introduced based on the particular physics and geometry of the given problem in order to introduce asymptotic forms and approximate solutions which may help convergence of the forward iterative scheme. The inversion result (11) then becomes a modified gradient solution [6] where the operator C may help to accelerate convergence and/or simplify computations in determining the model perturbation. Minimizing a weighted combination of data residual and solution size yields

$$\delta m_1 = \varepsilon \, C_m \, F^* \, C_d^{-1} \, \delta d \; , \tag{12}$$

as the first iteration result in a conjugate gradient least-squares inversion algorithm [6]. In (12), ε is a constant related to β times α in (11), and C_m and C_d are the least squares model and data covariance operators, respectively. The transpose of C in (11) is a model space dependent data weighting operator, incorporating a priori model and data covariance information as in (12). Single scattering described by the Born approximation can completely determine the model perturbation [7] so that in the inverse problem the transpose of the preconditioning operator must either help to describe the medium fully and use all the data for imaging, or it must become a model smoothing function to suppress the nonlinearities (multiple scattering) that are present in the data which will corrupt the image.

An example problem starts with a constant velocity background medium described by wavenumber k_o, and a change of variable $\phi(x,y)=\exp(ik_o x)P(x,y)$ transforms the 2-D scalar Helmholtz equation $\phi_{xx}+\phi_{yy}+k^2=0$ into the paraxial equation for P. Introducing a perturbation into the background slowness $\delta s=(s-s_o)$, where $\omega s_o = k_o$ (for monochromatic wave of frequency ω) gives the equation

$$P_{yy} + i \, 2\omega \, s_o \, P_x = -2 \, \omega^2 \, s_o \, \delta s \, P \; . \tag{13}$$

With the exact Green's function for the paraxial equation [2], the solution to (13) can be linearized

$$\frac{P}{P_o} - 1 = \frac{P_1}{P_o} = \int dx' \int dy' \; \kappa(x,y,\omega,x',y') \left(i \, \omega \, \delta s(x',y') \right) \; , \tag{14}$$

where the term $[i \, \omega \, P_o \, \kappa]$ is the Frechet derivative operator kernel, or sensitivity function for this problem. After a simple change of variables which translates the source to $(0,0)$ and rotates the coordinates so that the x'-axis is the linear ray

86

path between the source and the receiver a distance X away from the source, the function κ for this problem looks like

$$\kappa(X,\omega,x',y') = \frac{1}{H}\sqrt{\frac{1}{2\pi}}\ \exp\left\{i\left(\frac{(y')^2}{2H^2}-\frac{\pi}{4}\right)\right\}\ . \tag{15}$$

H has units of distance and is given by

$$H = \sqrt{\frac{(X-x')\ x'}{X\ \omega\ s_o}}\ . \tag{16}$$

The linearization given by (15) defines the Born approximation kernel for the paraxial wave equation, and it has been used to describe acoustic propagation in a cross-borehole seismic tomography imaging experiment [8]. For the following analytic results, a useful definite integral is

$$\int_{-\infty}^{\infty} e^{-ax^2}\ dx = \sqrt{\frac{\pi}{a}}\ , \tag{17}$$

from which one can derive

$$-\frac{\partial}{\partial a}\left(\int_{-\infty}^{\infty} e^{-ax^2}\ dx\right) = \int_{-\infty}^{\infty} x^2\ e^{-ax^2}\ dx = \frac{\sqrt{\pi}}{2}\ a^{-3/2}\ . \tag{18}$$

When there is no model perturbation, $\delta s(x',y')=0$ in (14) and $P_1=0$, implying no scattered component in the wavefield P. With a constant perturbation $\delta s(x',y')=\delta s_o$, using (17) yields

$$\frac{P_1}{P_o} = \int_0^X dx'\int_{-\infty}^{\infty} dy'\ \kappa(X,\omega,x',y')\left(i\ \omega\ \delta s_o\right) = i\ \omega\ \delta s_o\ X\ , \tag{19}$$

87

which is the correct phase change as seen from

$$\left(\frac{P}{P_o} - 1\right) = \left[\exp(i\,\omega\delta s_o X) - 1\right] = i\,\omega\,\delta s_o X + O\left(|\delta s_o|^2\right) . \tag{20}$$

Letting the perturbation be linear in y', $\delta s(x',y')=ay'$, (14) becomes

$$\frac{P_1}{P_o} = \int_0^X dx' \int_{-\infty}^{\infty} dy'\ \kappa(X,\omega,x',y')\ \left(i\,\omega\,ay'\right) = 0 , \tag{21}$$

since the integration over y' of the odd linear function against the even Gaussian function in (15) is zero. This result is correct for X small compared to the radius of curvature of the perturbed ray, which is consistent with the assumptions made in using the paraxial wave equation. Using (18), (14) can be evaluated for the harmonic oscillator problem where the slowness is a quadratic in y'

$$\frac{P_1}{P_o} = \int_0^X dx' \int_{-\infty}^{\infty} dy'\ \kappa(X,\omega,x',y')\ \left(i\,\omega\,\frac{a(y')^2}{2}\right) = -\frac{aX^2}{12s_o} . \tag{22}$$

Equation (22) describes the correct amplitude focusing (defocusing) given by the geometrical spreading equations [9] for a low (high) velocity channel in which the second derivative of the slowness is negative (positive).

Using (10) to define the operator C with kernel σ gives a new function κ_s for use in (14)

$$\kappa_s(X,\omega,x',y') = \int dx'' \int dy''\ \kappa(X,\omega,x'',y'')\ \sigma(x'',y'' \mid x',y') . \tag{23}$$

The function σ can be a function of frequency and source/receiver geometry as well as that of the scattering points within the medium. Appropriate functions for σ can be determined under the constraint that the ray theoretical results of (19), (21) and (22) be retained. For the class of functions given by

$$\sigma(x'',y'' \mid x',y') \;=\; \delta(x''-x') \sqrt{\frac{1}{2\pi}} \frac{1}{F} \exp\left(\frac{y'}{FH} y'' - i\frac{\pi}{4} \right) \qquad (24)$$

the new kernel κ_s is exactly like in (15) except F replaces H. In order to retain
the ray-theoretical results, F must equal H itself or it must satisfy

$$F = \sqrt{\frac{x'}{3\omega\, s_o}} \;,\quad F = \sqrt{\frac{(X-x')}{3\omega\, s_o}} \quad\text{or}\quad F = \sqrt{\frac{X}{6\omega\, s_o}} \qquad (25)$$

which describe near-source, near-receiver and far-field x'-dependence in the
sensitivity kernel, respectively. Using the far-field kernel derived through (23),
(24) and the third term in (25), the square-root singularities at the source and
receiver are eliminated so that the convergence criterion in (9) is easier to
achieve. Another class of functions for σ is

$$\sigma(x'',y'' \mid x',y') \;=\; \delta(x''-x') \sqrt{\frac{1}{2}} \frac{1}{F} \exp\left[- \left(\frac{H}{F}\frac{y'}{y''} \right)^2 - i\frac{\pi}{4} \right] , \qquad (26)$$

where F is as defined for (24). Equations (24) - (26) can be determined with the
aid of the two useful definite integrals

$$\int_{-\infty}^{\infty} \exp\left(ax^2 + bx \right) \, dx \;=\; \sqrt{\frac{\pi}{-a}} \, \exp\left(\frac{-b^2}{4a} \right) \qquad (27)$$

and

$$\int_{0}^{\infty} \exp\left(-\frac{a}{x^2} - bx^2 \right) \, dx \;=\; \sqrt{\frac{\pi}{4b}} \, \exp\left(-2\sqrt{ab} \right) . \qquad (28)$$

The new kernel obtained using (26) in (23) with (15) is

89

$$\kappa_s(X,\omega,x',y') = \frac{1}{F}\sqrt{\frac{1}{2}} \ \exp(-\frac{|y'|}{F}) \ \exp\left[i\left(\frac{|y'|}{F} - \frac{\pi}{4}\right) \right] \ .$$
(29)

Equation (29) describes a sensitivity function having ray-perpendicular amplitude variation that scales as the Fresnel zone size, with the sensitivity damping out away from the classical ray. The square-root singularities of the sensitivity function defined by (29) can be controlled by the choice of F as with the kernel obtained by (24). Also, the practical numerical integration range over y' can be reduced due to the exponential decay term {exp(-|y'|/F)}, so that the convergence criterion (9) can be achieved.

The functions defined by (24) and (26) are model space smoothing operators in the inversion solution of (11). Using (23) and interchanging the order of integrations, equation (14) shows the true sensitivity function given by κ acting on a smoothed version of the model parameter

$$\hat{\delta s}(x'',y'') = \int_0^X dx' \int_{-\infty}^{\infty} dy' \ \sigma(x'',y'' \mid x',y') \ \delta s(x',y') \ .$$
(30)

The inverse problem is transformed into one of recovering the smoothed model parameter defined by (30). After the rotation of coordinates to get equation (15), the three sets of rectangular coordinates used in equations (14) and (30) are coincident, and (30) gives the relationship between the amplitude of the model parameter field in the two image-coordinate systems. Figure 1 shows the simple geometry for a source and a receiver a distance of .5 km apart. The examples in Figures 2 and 3 show 2-D plots of smoothed slowness models for a 50Hz monochromatic wave and a background velocity of 2 km/sec giving a Fresnel zone scaling halfway between the source and receiver of about 0.1 km. Starting from an initial model perturbation of a constant value along a line parallel to the ray path (x-axis) in (30), the function (26) gives a more slowly varying model parameter distribution that can be recovered in an imaging experiment.

Figure 2 with F=H and Figure 3 with F the far-field term of (25) are similar, each showing new scattering amplitudes spread to both sides of the classical ray, incorporating the uncertainty in phase accumulation relative to the stationary phase path. When a scattering point is along the ray path y'=y=0 (Figures 2a and 3a), the slowness amplitudes are smeared by the smoothed sensitivity kernel into a constant value across the transverse coordinate given by the scaling function F. With a scattering position further from the ray path (Figures 2 and 3,b-d), the scattering amplitudes are smeared from the original scattering location with increasing amplitude away from the ray path. Correct image reconstruction is attainable with sufficient ray coverage, where stationary points constructively interfere using the phase accumulation of the sensitivity kernel for each ray path relative to the true scatterer location. In the case of poor ray coverage, smoothing can become important for creating a physically reasonable image [8]. Smooth models also help iterative inversion methods to be more stable. When a smoothed image is desired, defining the smoothing function through the preconditioning operator in the forward theory may be advantageous in allowing for approximate solutions to directly incorporate a priori information into the inversion.

REFERENCES

[1] Kleinman R E and Van den Berg P M 1991 Iterative methods for solving integral equations *Radio Science* 26 175-81

[2] Mathews J and Walker R L 1970 *Mathematical methods of physics* (Menlo Park, CA: Benjamin/Cummings Publishing Co.)

[3] Murch R D 1992 An extended Born approximation *Inverse Problems* 8 L5-11.

[4] Cardimona S 1991 Waveform inversion and digital filter theory *Geophysics* 56 534-6

[5] Petryshyn W V 1963 On a general iterative method for the approximate solution of linear operator equations *Math Comp.* 17 1-10

[6] Tarantola A 1984 Linearized inversion of seismic reflection data *Geophys.Prosp.* 32 998-1015

[7] Snieder R 1990 The role of the Born approximation in nonlinear inversion *Inverse Problems* 6 247-66

[8] Cardimona S and Garmany J 1993 Smoothing operators for waveform tomographic imaging *Geophysics* (submitted)

[9] Cerveny V and Hron F 1980 The ray series method and dynamic ray tracing system for three-dimensional inhomogeneous media *Bull. seism. Soc. Am.* 70 47-77

Figure Captions

Fig. 1. Geometry in ray-centered coordinates for 2-D slowness model around a straight ray connecting source and receiver positions. Distances in km.

Fig. 2. Ray-centered geometry of Figure (1). Slowness models given by the smoothing function in equation (26) with F=H. Original slowness perturbation was a line at y equal to (a) 0.00, (b) 0.02, (c) 0.05, (d) 0.08 km.

Fig. 3. Ray-centered geometry of Figure (1). Slowness models given by the smoothing function in equation (26) with F equal to the far-field scaling term of equation (25). Original slowness perturbation was a line at y equal to (a) 0.00, (b) 0.02, (c) 0.05, (d) 0.08 km.
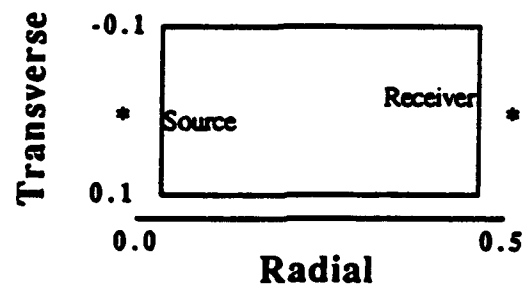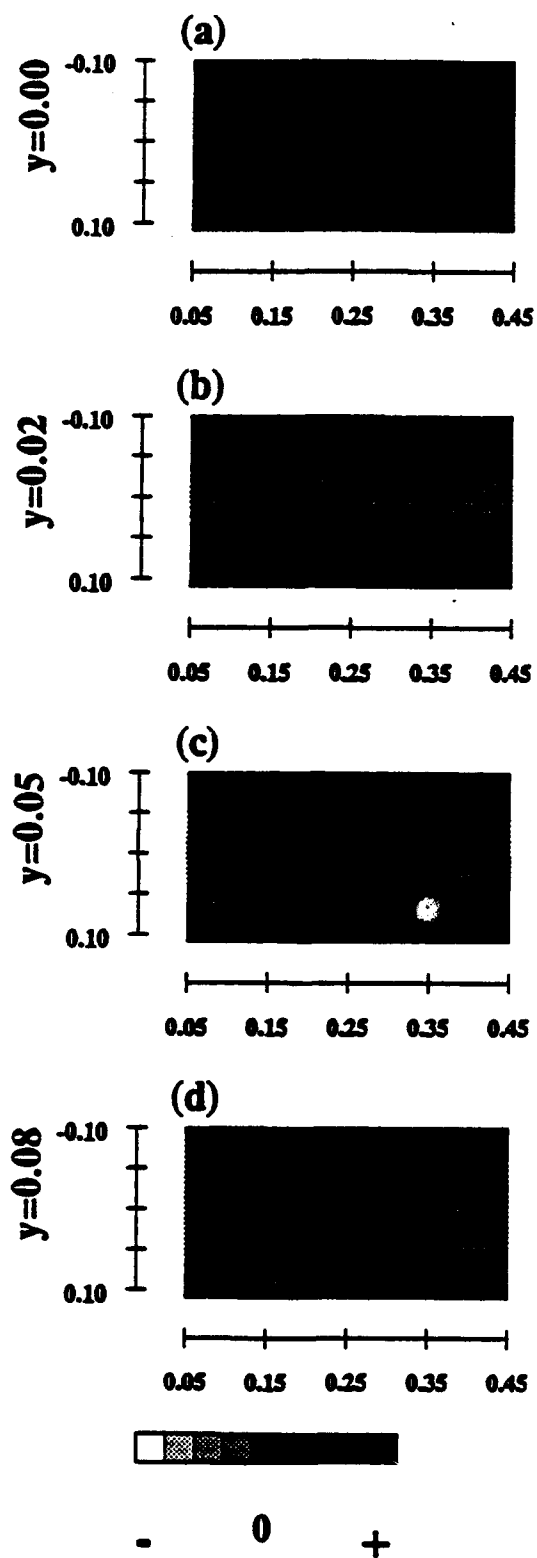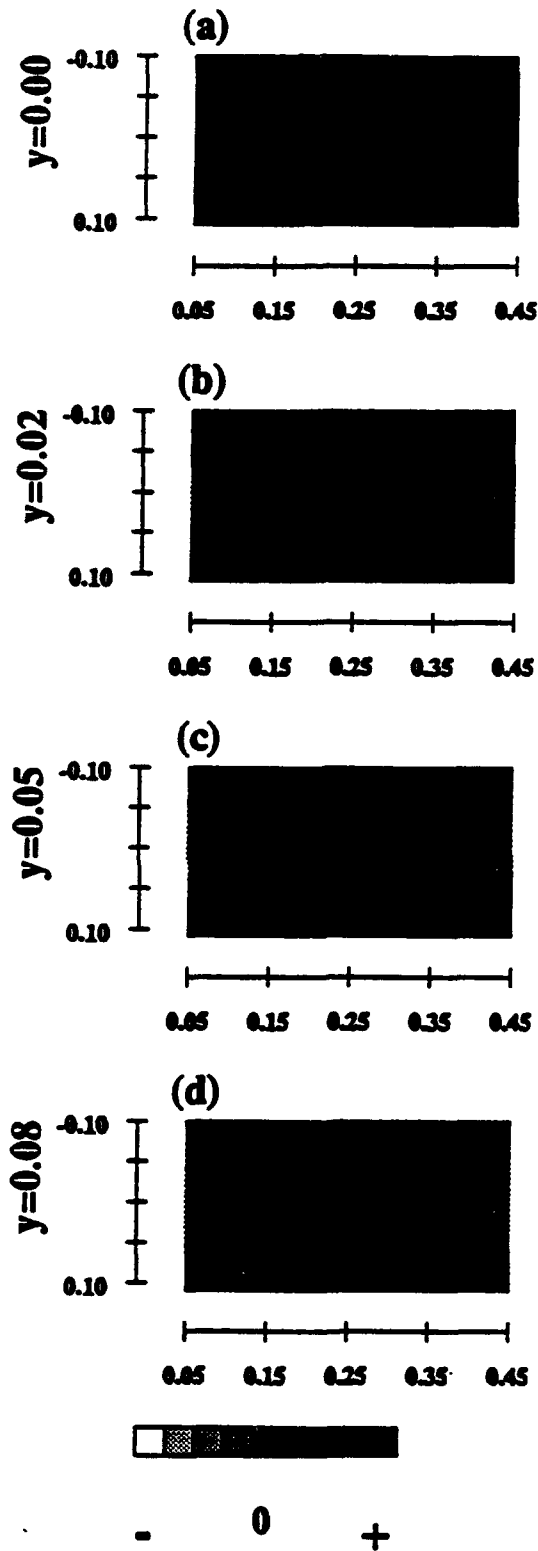
FIGURE 1

FIGURE 2

FIGURE 3

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

MAPPING THE 660-KM DISCONTINUITY THROUGH A DESCENDING SLAB

Prepared by: Dr. Steve Cardimona

Research Location: Phillips Laboratory,
                                     Geophysics Directorate
                                     Hanscom AFB, MA 01731

AF Research Colleague: Anton Dainty

Contract Number: F19628-86-C-0224

Mapping the 660-km discontinuity through a descending slab

Steve Cardimona* and Jan Garmany**

*Phillips Laboratory/GPEH, Hanscom AFB, MA 01731-3010

**Institute for Geophysics, Austin, TX 78759

Long period seismic data exhibit coherent phases from the discontinuity at 660 km depth within the earth ("the 660") [1,2,3] and these data suggest that topography of the boundary correlates with regional tectonic features [4]; but these data are limited to a lateral resolution of more than 1000 km. The analysis of the moveout of short period S-to-P conversions at the 660 (S660P) requires that the discontinuity near subduction zones be about 25-30 km deeper than the global average [5,6]; however, the small-scale lateral variation in the vicinity of the slab has yet to be imaged. A model for the 660 which best explains the coherent scattering in the short period data of this study is one in which the discontinuity is due to an endothermic phase transition in mantle minerals, causing a localized depression of the boundary within the cold slab. The depression is measured at 41 km, indicating the temperature difference between the mantle and the slab interior is 820 K.

The 660 discontinuity is primarily due to the phase transition of γ-olivine (spinel) to perovskite + MgO ("spinel to perovskite transition"), since a chemical boundary in this region would be fairly invisible to seismic data and would not necessarily be associated with the 660 km depth [7]. Recent results of seismic data analyses are consistent with this interpretation [2,5,8], but have yet to resolve the lateral variation of the boundary in the vicinity of the slab which would give important information about the temperature structure in this region. This study examines short period, vertical component records from North America for deep earthquakes south of Fiji Island. The data sample a region around 300 km about

the slab in the direction perpendicular to the trench (Fig. 1). The coda of the first arrival (P wave) can be associated with a model of near-source structural scattering, including compressional and shear energy reflected from above the earthquake, and shear wave energy converted to compressional energy below the earthquake. The different moveout of each phase allows the analyses to be distinct, so that stacking for one type of scattered phase will be uncorrupted by other coherent phases.

Associating the P wave coda with shear to compressional wave (S to P) conversion below the earthquake (called SdP, where 'd' is the depth of conversion), we interpret the scattering using waveform inverse theory [9]. With a one-dimensional linear model , a gradient solution to the inverse problem reduces to application of a matched filter [10], and a velocity model (PREM [11]) gives the mapping from time to depth. For S to P conversions, the time after P at which scattered energy arrives is given by an integral over depth of the difference between the S wave vertical slowness and that of the P wave. Travel-time differences relative to the first arrival are station independent, and are associated with structure relative to the source position. This defines a delay-time analysis such that, when the data are stacked, the coherent near source phases are enhanced over any ray path and near-receiver scattering. The process is similar to a slant-stack [5,6], however the near-source sampling and source-receiver geometry are used directly to create an image of the structure in the vicinity of the slab. Near-P energy related to individual source characteristics is not coherent when stacking records from different earthquakes, so that misinterpreting this energy as due to upper mantle scattering is minimized.

Useful data records are those for which we were able to pick good arrival times for the P wave and for the pP phase defined by the sea-floor reflection. The (pP-P) travel time difference gives the depth for each earthquake relative to PREM, and this serves to constrain the structure analysis. The earthquake sources are in general down-dip compressional [12], resulting in both shear and compressional energy in the azimuth to North America. Waveforms from a variety of deep earthquakes in the Tonga subduction zone have been successfully described by a single simple source time wavelet [13], and cross-correlating a minimum phase source wavelet estimate [14] for each corresponding trace of this study makes the data

records close to zero phase. Stacking the data with this minimum phase assumption is justified since introducing random rotation into each record before stack shows the results are unaffected until the random phase gets as large as 90 degrees. The total energy of each record is used to normalize that record's amplitude relative to those of other earthquakes. With the amplitude and phase normalized, all the records can be stacked together. The choice of background velocity model does not affect the inverse problem, since most velocity models agree to within 2%, and the analysis described here is insensitive to this amount of perturbation.

Figure 2a displays a record section of the 1-D depth-mapped data, with respect to earthquake depth. Limited stacking of data is required to reduce the noise and enhance any signal, and each vertical trace in Figure 2a is the sum of 8 nearest neighbor data traces, with the single trace stack of all the data on the right. Figure 2b is a section where each vertical trace is a stack including two more data records than the trace to its left, with the single trace stack of all the data on the far right. The clearest stacking result comes from S to P conversion at 675 km below the sea floor, and is consistent with other observations in this region [2,5,6], relative to PREM. The amplitude of the event in the one-dimensional stack (Fig. 2) gives an estimate of the S660P transmission coefficient of about 3% relative to incident S wave energy, which is consistent with PREM for the angles of incidence of the data set, and is only about 1% higher than for an earth model with slightly smaller contrasts in P and S wave velocities and density than with PREM. Another feature at around 715 km shows up in the data stacking section of Figure 2b, and may be due to conversion at a deeper interface, although petrological data does not support this. An alternate interpretation is that the scattering at 715 km is due to the spinel to perovskite transition depressed within the slab [5]. The phase transition has a negative Clapeyron slope which describes the differential change in pressure with respect to a change in temperature, moving the phase boundary to greater depth within the cold slab relative to the surrounding mantle. The band-limited seismic data of this study illuminate the interface with first Fresnel zone widths between 30 and 60 km. When the 660 is perturbed, there may be more than one stationary point giving rise to coherent scattering.

The model parameters of slab width and change of depth within the slab are

introduced to describe the perturbed interface of the 660, and the data are inverted by iterative forward modeling using Fresnel-Kirchhoff diffraction theory (Fig. 3) to model the phase variation in the stacked data. The third dimension is suppressed, assuming high correlation in the along-trench direction. The simple model of the interior of the slab is a first approximation, matching the expected variation at the leading edge of the slab where the temperature change is most dramatic, although the temperature structure of the under side of the slab would be more gradual (Fig. 1). The wavelength resolution in depth is evident from the del-Z variation in Figure 3, but there is a well resolved peak at 41 km. The slab width parameter is more poorly defined than the change in depth. This is due in part to the imperfect representation of the phase boundary relative to the temperature structure (Fig. 1), and is sensitive to the absolute amplitudes of the data and the synthetics. The result is a broad variation in slab, approaching a constant value at a width of 0.0 km (no perturbation) and with a maximum at 50 km.

In Figure 4 are images created using the source-receiver geometry to bin and stack the data into a trench-perpendicular image plane. Comparing the 1-D and 2-D imaging results for the data (Fig. 4a) and the synthetics (Fig. 4b) created for the preferred model of the 660, with a 50 km wide slab (in the direction transverse to the trench axis) and a 41 km depression of the boundary within the slab, shows that diffraction theory describes the qualitative features of amplitude and phase variation seen in both the 1-D and 2-D data results (Fig. 4a), and is superior to a two-interface model (Fig. 4d). The diffractions imaged with the synthetics coincide with the bright features in the data image, displaced from the model due to the experimental geometry (Fig. 4c).

A measurement of the depression in the 660 gives a direct indication of the temperature difference between the cold slab and the surrounding mantle. The increase in pressure with change in depth gives the Clapeyron slope in terms of the ratio of depth change to temperature change. A preferred value for the Clapeyron slope is $\gamma = -2.0$ MPa/K (ref. 15). Taking the acceleration of gravity to be 10 m/s^2, estimating the density at 4000 kg/m^3, and using the measurement of the change in depth of $\delta z = 41$ from this study, the temperature difference between the slab and the mantle is calculated to be -820 K. When the spinel to perovskite phase

transition is introduced into a realistic mantle model with this temperature contrast, whole mantle convection patterns can be disrupted by the phase transition resulting in intermittent mixing between the upper and lower mantle [15]. Recent tomographic results are consistent with this, finding evidence for slab penetration into the lower mantle in some places and slab deflection in others [16].

References

1. Shearer, P. M. *Nature*, 344, 121-126 (1990).

2. Bock, G. and Kind, R. *Geophys. J. Int.*, 107, 117-129 (1991).

3. Revenaugh, J., and Jordan, T. H. *J. Geophys. Res.*, 94, 5787-5813 (1989).

4. Shearer, P. M., and Masters, T. G. *Nature*, 355, 791-796 (1992).

5. Richards, M. A., and Wicks, C. W. *Geophys. J. Int.*, 101, 1-35 (1990).

6. Vidale, J. E., and Benz, H. M. *Nature*, 356, 678-683 (1992).

7. Jeanloz, R. *Geophys. Res. Lett.*, 18, 1743-1746 (1991).

8. Wicks, C. W. and Richards, M. A. *Geophys. J. Int.*, 107, 279-290 (1991).

9. Tarantola, A. *Geophys. Prosp.*, 32, 998-1015 (1984).

10. Cardimona, S. *Geophysics*, 56, 534-536 (1991).

11. Dziewonski, A. M., and Anderson, D. L. *Phys. Earth Planet. Inter.*, 25, 297-356 (1981).

12. Giardini, D., and Woodhouse, J. H. *Nature*, 307, 505-509 (1984).

13. Pennington, W. D., and Isacks, B. L. *Geophys. J. R. astr. Soc.*, 56, 19-40 (1979).

14. Yilmaz, O. *Seismic Data Processing*, 500-502, (Society of Exploration Geophysicists, Tulsa, OK, 1987).

15. Machetel, P., and Weber, P. *Nature*, 350, 55-57 (1991).

16. van der Hilst, R., Engdahl, R., Spakman, W., and Nolet, G. *Nature*, 353, 37-43 (1991).

Figure Captions

FIG 1: The data of this study consists of 129 North American Global Digital Seismograph Network records from 46 events South of Fiji Island with body-wave magnitude greater than or equal to 5.5. The azimuth to the North American receivers is a few degrees East of the trench axis. The crosses show the points at 675 km depth sampled by the data set with respect to the geometry of the trench and subduction zone. Inset shows the leading edge of the slab at 675 km depth, with the trace of the expected phase transition boundary and the model for the boundary used in this study. Data and synthetics are used to create an image of structure perpendicular to the trench axis.

FIG 2: Data are low-pass filtered at .8 Hz, and the amplitude and phase are equalized. (a) Data record section after applying the SdP moveout analysis plotted relative to the depth of the earthquake, with each vertical trace the sum of 8 data records. Associating the first arrival with the source depth, the energy after the first arrival is mapped to scattering from positions below the earthquake. Zero depth in the image is defined to be associated with the sea floor reflection, at 3 km depth in PREM. The impulsive, high amplitude P arrival is suppressed with a 120 km cosine taper (about 15 seconds in time) on the depth-mapped data records before stack. The single trace stack of all the data is on the right. (b) From left to right, each vertical trace has two more data records stacked together than the previous trace, up to the single trace stack of all the data at the far right. The S675P phase is seen well in both the one-dimensional stack and the grey-scale plot of the depth-mapped data. In the record section, there is another structural feature at around 715 km depth interpreted as a diffraction from the spinel to perovskite transition within the slab.

FIG 3: Objective functions for the model space defined by slab width and depth of interface (del-Z) within the slab relative to the 675 km depth in the surrounding mantle. (a) Correlation coefficient. (b) Error energy. Fresnel-Kirchhoff diffraction theory and a zero phase modeling wavelet are used to create noise-free synthetic seismograms for a single interface that is depressed within the slab, assuming high correlation in the along-trench direction. Synthetics were calculated for 121 models, varying the slab width from 0.0 km (no slab expression) to 100 km, and the depression of the 675 km seismic boundary within the slab between 30 and 60 km. The correlation coefficient statistic in (a) shows the energy that is in phase between the data and the synthetic. The normalized squared-error (total error energy) objective function in (b) is defined so that low error shows up darkest. The resolution of the slab width parameter given by the correlation statistic is a bit less sensitive to the absolute amplitudes. The best model is described by a 50 km wide slab with a 41 km depression of the phase boundary within the slab.

FIG 4: (a) 2-D structural imaging of data in Fig. 2. The source-receiver geometry is used to bin and stack subsets of the individual depth records into the two-dimensional image perpendicular to the trench axis, with a horizontal coordinate in East longitude and depth in km. Nearest neighbor horizontal averaging is applied to get the final 2-D image, and the single trace stack is on the right. The S675P phase is imaged well in both the one and two-dimensional plots. In the 2-D image, there is another structural feature at around 715 km depth interpreted as a diffraction from the spinel to perovskite transition within the slab. (b) Imaging results for the synthetics created for the best fitting model from Figure 3, with a slab width of 50 km and del-Z equal to 41 km. (c) The trace of the best model where it intersects the image plane, plotted with the data image. (d) Imaging results for synthetics from a two-interface model. A zero phase modeling wavelet was used to calculate ray-theory synthetics created to model the amplitude and phase variation of the data records in order to reproduce the strong features of the actual stacked data image in the vicinity of the 660 km discontinuity. The one (perturbed) interface model is superior to the two interface model.
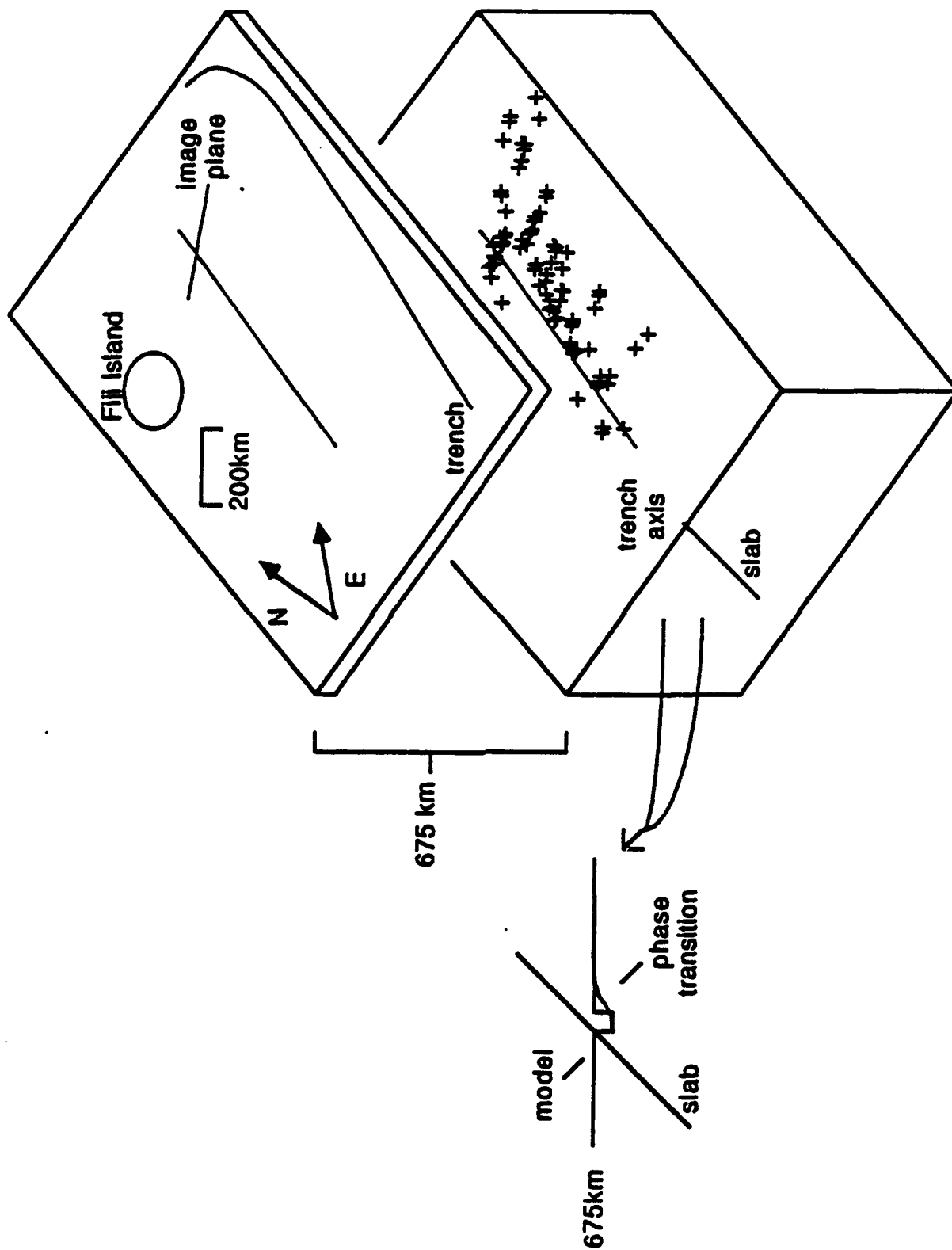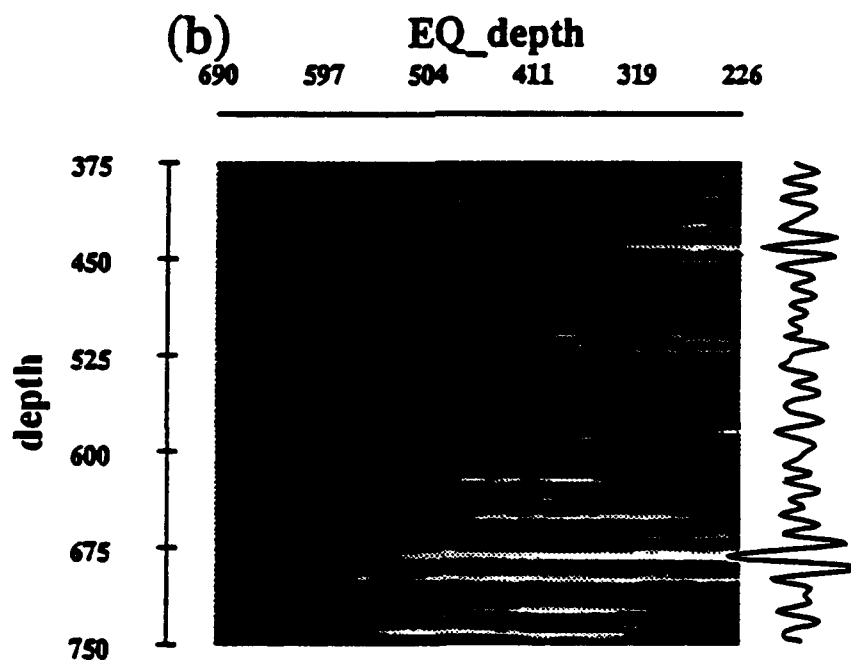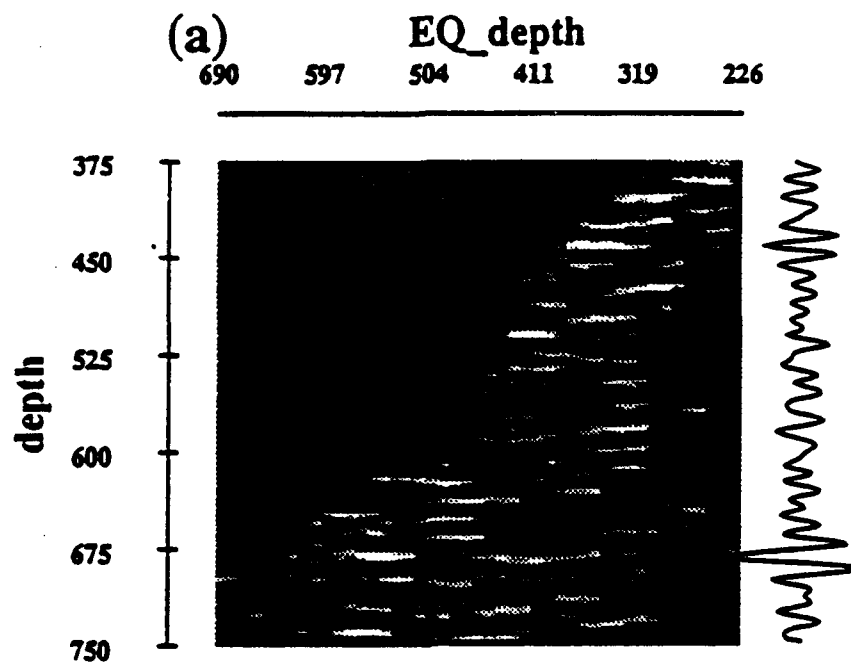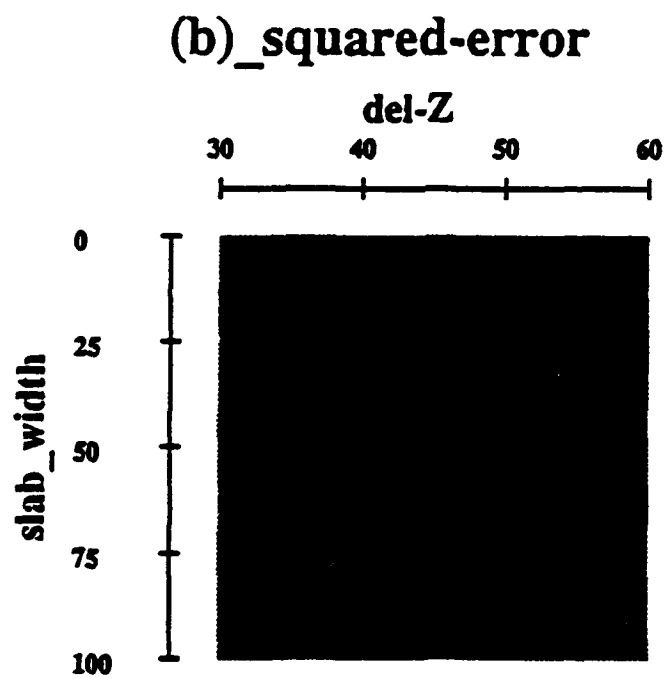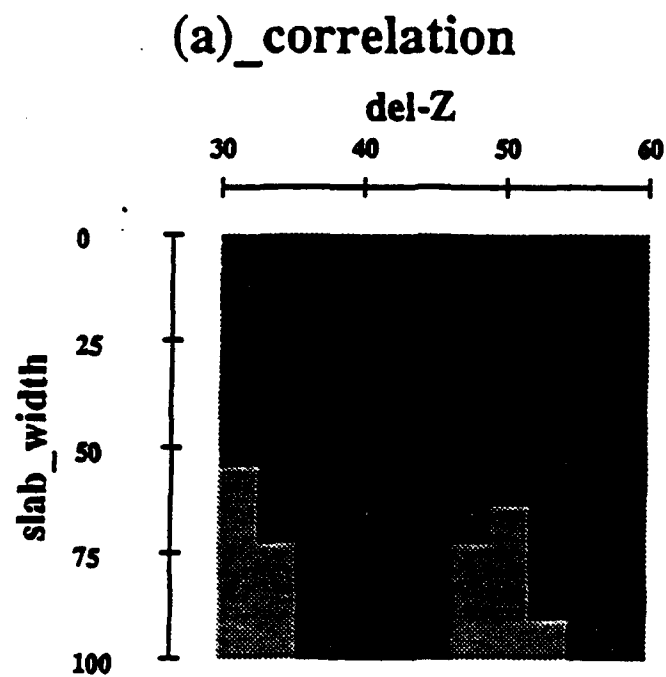
FIGURE 1

106

FIGURE 2

(a)_correlation

(b)_squared-error

FIGURE 3

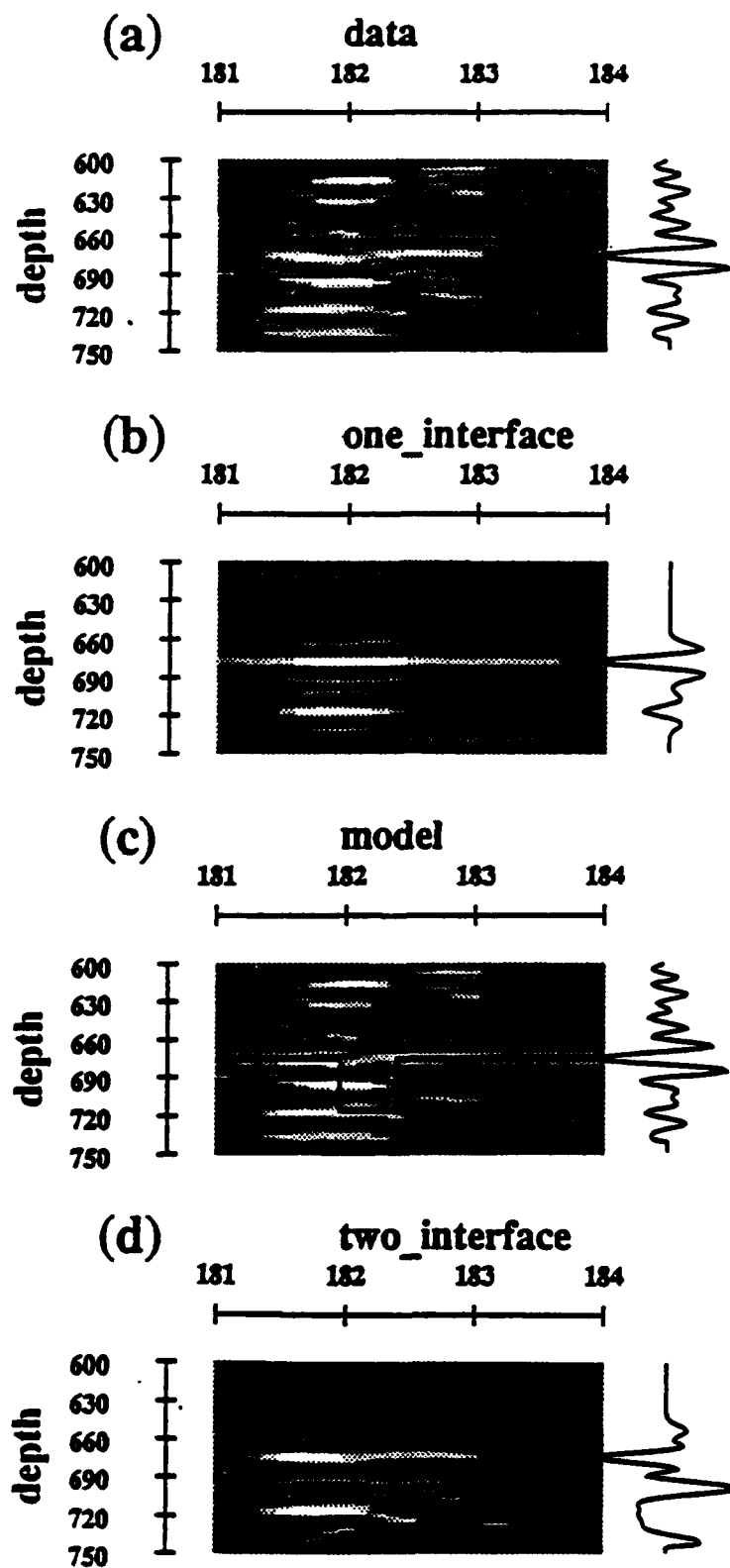FIGURE 4

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

WORKS ON TROPICAL CYCLONE DEVELOPMENT AND WORKS ON CLIMATE CHANGE

Prepared by: Dr. Mark Handel

Research Location: Phillips Laboratory,
                                      Geophysics Directorate
                                      Hanscom AFB, MA 01731

AF Research Colleague: J. William Snow

Contract Number: F19628-86-C-0224

This final report of Works on Tropical Cyclone Development and Works on Climate Change by Mark David Handel as a SCEEE Geophysics Scholar is divided into four parts. It covers the period 1 June 1991 to 29 May 1992.

Part I. Tropical Cyclone Intensification:
Ekman Pumping and the Formation of an Eye.

The single paper in this part of the report was also submitted to the *Journal of the Atmospheric Sciences* under the title "Tropical cyclone intensification. Part III: Ekman pumping and the formation of an eye". The work going into this report was done almost entirely during the period of the contract.

Part II. On the Energetics of Asymmetric Tropical Cyclone Outflows.

An abstract based on this work was published as "On the asymmetry of tropical cyclone outflows" in *EOS, Transactions of the American Geophysical Union* 73 (14, Spring Meeting supplement), p. 73 (1992); and a talk was given at the 1992 Spring Meeting of the American Geophysical Union. An earlier talk on this topic was given at the Satellite Meteorology Branch of Phillips Laboratory. The work going into this report was done almost entirely during the period of the contract.

Part III. Other Tropical Cyclone Papers.

The two papers in this part of the report were submitted to the *Journal of the Atmospheric Sciences* under the titles "Tropical cyclone intensification. Part I: Impediment to linear growth" and "Tropical cyclone intensification. Part II: Growth from finite amplitude disturbances". An extended abstract based on this work was published as "Subcritical tropical cyclone intensification" in the preprint volume of the 8th Conference on Atmospheric and Oceanic Waves and Stability, American Meteorological Society, Boston, Massachusetts, pp. 315–319 (1991); and a talk was given at that conference. The work going into this report was done largely prior to the period of the contract, though the papers were written during the period of the contract.

Part IV. Climate Change Papers.

The first paper in this Part will be published as "Reflections on more than a century of climate change research in *Climatic Change* 21, pp. 91–96 (1992), authored by Mark

David Handel and James S. Risbey (of the Massachusetts Institute of Technology). It was the result of Dr. Handel serving as guest editor of an issue of *Climatic Change*. The piece was written entirely during the period covered by the contract.

The second paper in this Part will be published in slightly amended form as "An annotated bibliography on the greeenhouse effect and climate change" in *Climatic Change* 21, pp. 97–255 (1992), authored by Mark David Handel and James S. Risbey (of the Massachusetts Insitute of Technology). The paper was significantly expanded during the period covered by the contract from work performed earlier.

The last two papers in this Part were published as "Do we know what difference a delay makes?" in *EOS, Transactions of the American Geophysical Union* 72 (53), pp. 596–597 (1991); and "Should we delay responses to the greenhouse issue?" in *EOS, Transactions of the American Geophysical Union* 72 (53), pp. 593 (1991). Both papers were authored by James S. Risbey, Mark David Handel, and Peter H. Stone (both Risbey and Stone of the Massachusetts Institute of Technology) and were part of an exchange with Michael E. Schlesinger and Xingjian Jiang (included for completeness). The first of the papers was prepared prior to the period covered by the contract and revised during the period of the contract. The second paper was prepared during the period covered by the contract.

1991–1992 SCEEE Geophysics Scholar Program

Sponsored by

Department of the Air Force
Phillips Laboratory / Geophysics Directorate

Conducted by the

Southeastern Center for
Electrical Engineering Education

Final Report

Part I. Tropical Cyclone Intensification:
Ekman Pumping and the Formation of an Eye

Prepared by: Dr. Mark David Handel
Research Location: Phillips Laboratory / Geophysics Directorate
Phillips Laboratory Research Contact: Dr. J. William Snow
Date: May 29, 1992
Contract Number: F19628-86-C-0224

ABSTRACT

An open, cloud-free, low wind, region, known as an *eye*, usually forms in the center of intense tropical cyclones. This region is surrounded by an annular *eyewall*, containing a storm's strongest winds and most intense convection. This note expands on earlier suggestions that the formation of the eyewall, and hence also the eye, is largely controlled by surface friction. CISK (Conditional Instability of the Second Kind) dynamics are assumed, so deep convection is assumed to be initiated by Ekman pumping. Simple theoretical examination are made of Ekman pumping with different stress laws. Changes in intensity are shown to result in changes in the radial profile of Ekman pumping, such that weak disturbances have strong central convection, while in more intense disturbances the convection forms a ring.

When that Cloud begins to move apace, you may expect the wind pre∫ently. It comes on fierce, and blows very violent at N.E. 12 hours more or le∫s. It is also commonly accompanied with terrible claps of Thunder, large and frequent fla∫hes of Lightning, and exce∫∫ife hard rain. When the wind begins to abate it dies away suddenly, and falling flat calm, it continues so an hour, more or le∫s: then the wind comes about to the S.W. and it blows and rains as fierce from thence, as it did before at N.E. and as long. (Dampier 1699)

1. Introduction

A cloud-free, nearly calm, central *eye* is one of the more dramatic features of a tropical cyclone. The eye is closely surrounded by the most intense convection and strongest winds of the storm in an annular region known as the *eyewall*. Similar structures are also observed in intense nearly axisymmetric storms of the middle and upper latitudes (Gyakum 1983, Rasmussen 1979). Ooyama (1969) was perhaps the first to attribute the development of an eyewall, and its accompanying eye, to the radial profile of frictionally forced vertical velocity out of the boundary layer. He reached this conclusion while interpreting results from his numerical model. We briefly examine why weak vortices tend to have maximum convection located near a central vorticity maximum, while stronger ones develop an eyewall.

Inherent throughout this analysis is the assumption that CISK (Conditional Instability of the Second Kind) processes are at work, with deep cumulus convection initiated in regions of positive Ekman pumping. With this underlying assumption, we will examine spatial profiles of Ekman pumping for different drag laws and symmetries. Differences in these profiles will then be applied to an analysis of tropical cyclone development. A brief discussion of some considerations in choosing a surface stress parameterizations is also included.

1.1. Explanations of Eye Formation

A review of theories of eye formation can be found in Anthes 1982. There are two somewhat separate issues in the formation of an eye. One is why no convection originates in the center of a mature tropical cyclone. The second is why there is descent in the upper levels of cyclone near the central axis. These are not totally unrelated, since descent and warming in the center will eventually stabilize the column and inhibit vertical convection.

Several works on eye formation and maintenance have concentrated on explaining the descent of high entropy dry air into the eye. Older works (e.g. Byers 1944) made attempts to attribute the descent to centrifugal forces flinging air from the central core, requiring air from aloft to replace it. Malkus (1958) suggested that descent in the upper part of the eye was driven by the disequilibrium of supergradient winds (winds greater than those that maintain a balance between the pressure gradient with the Coriolis and centrifugal forces) and momentum mixing. A closely related argument for descent driven by momentum mixing was made independently by Kuo (1959).

Though Palmén (1948) attributed the high eye temperatures to "the tendency to establish a combined hydrostatic and geostrophic-cyclostrophic balance", W?loughby (1979) was the first to note that the descent aloft could purely be an adjustment response to the radial gradients of the heating profile and that momentum diffusion was not necessary. More detailed discussions of similar processes were provided by Smith (1980). Though they did not concentrate on eye formation, Schubert et al. (1980) noted that forcing of a cyclone through the vorticity field by cloud clusters was more efficient than forcing by direct heating. Geostrophic adjustment to the velocity field of a vortex then requires descent in the center. The sinking of high entropy dry air is also evident in numerical models (e.g. Kurihara and Bender 1982).

The heating of ascending parcels cannot provide a temperature perturbation as large as is observed in the upper troposphere during intense tropical cyclones (about 15 K according to Hawkins and Rubsam 1968). Further, the accompanying surface pressure perturbation is too large to be created by ascent of parcels of the greatest available entropy from the surface, even after saturation. However, descent of high entropy air from above is consistent with sounding observations (Franklin *et al.* 1988) and weak descent has also been observed directly (Jorgensen 1984). If a cyclone is close to neutral for slantwise convection, as suggested by Emanuel (1986), there would be descent aloft between parcel paths originating near the cyclone center at the surface and slanting outwards. So though sinking air in the upper troposphere, and its accompanying stabilization of the column, must be an integral part of the eye formation process, it cannot explain the central region of very warm moist surface air where no convection originates.

Other works have concentrated on the locations of regions of ascent in the boundary layer creating the eyewall. Syōno (1951), Abdullah (1953), and Kuo (1959) all presented arguments that parcels originating far from the central axis could only reach some minimum radius, and then must head upwards. These arguments, to first approximation, assumed conservation of angular momentum, noted that the kinetic energy of a parcel conserving angular momentum had to continually increase as the radius of the parcel decreased, and used some limit on the energy or pressure gradient to then provide a minimum radius a parcel could reach. These works ignored the fact that boundary layer inflow, especially at steady state, is largely *driven* by momentum dissipation, so anugular momentum is significantly not conserved within the boundary layer. So, though these works attempted to address the cause for the annular region of ascent to be at some distance from the center, they ignored the primary reason for the existence of any boundary layer inflow.

As already noted, Ooyama (1969) suggested that the radial profile of Ekman

pumping caused the formation of an eyewall. In an analytic study of Ekman layers with different types of boundary conditions, Eliassen (1971) showed that in a vortex with Taylor's slip condition for a turbulent boundary layer, the forced vertical velocity at the center vanished and the maximum positive pumping was found in a ring away from the center axis. These results were extended by the numerical calculations of Eliassen and Lystad (1977). In contrast, for a vortex in a viscous fluid with a no-slip condition the Ekman pumping approaches a constant at the center axis (Greenspan and Howard 1963). With these analytic results, Eliassen suggested that the suppression of ascending motions near the axis of a vortex with a turbulent Ekman layer might control the location of convection and lead to the creation of a tropical cyclone eye. Though Eliassen did not go on to suggest that solutions that approach constant Ekman pumping in the center might be relevant to weak disturbances, which have not yet developed an eye, we will make such a connection here. The numerical simulations of Yamasaki (1977) showed the necessity of surface friction for the development of an eye. Yamasaki also suggested that surface friction is unimportant in the early development of a tropical cyclone, though perhaps this is because his model was axisymmetric and included only a quadratic drag law.

2. Profiles of Ekman Pumping

The model used here is the same as that originally developed by Ooyama (1969) and presented in Part I (Handel 1991a), appropriate for examining a simplified tropical cyclone. It is set in cylindrical coordinates (r, z), where r and z are the radius and height. The disturbance is centered and fixed on the origin of the coordinate system. All fields are assumed to be axisymmetric. The coordinate system is rotating and the Coriolis parameter, f, is constant. The cyclonic tangential, or azimuthal, velocity is labeled v_0 and the outward radial velocity is u_0. The continuity equation for the

118

constant density, ρ_0, boundary layer of thickness h_0 is

$$\frac{\partial h_0}{\partial t} = \frac{\partial \psi_0}{r\, \partial r} - w \ , \tag{2.1}$$

where w is the vertical velocity at the top of the boundary layer and $\psi_0 \equiv -h_0 u_0 r$ is the inward mass flux. The equation for the vertical component of the angular momentum, $M_0 \equiv v_0 r + f r^2/2$, is

$$\frac{\partial}{\partial t}(h_0 M_0) = \frac{\partial}{r\, \partial r}(\psi_0 M_0) - Z_{01} + Z_s \ , \tag{2.2}$$

where Z_s is the vertical flux of angular momentum at the surface and Z_{01} is the vertical flux of angular momentum between the boundary layer and the layer above it. In the boundary layer, the lateral momentum diffusion is assumed to be much smaller than the vertical fluxes and is ignored. The vertical flux at the surface is

$$Z_s = -\tau_s r/\rho_0 \ , \qquad \tau_s = \rho_0 \left[k_s v_0 + C_D (u_0^2 + v_0^2)^{1/2} v_0 \right] \ , \tag{2.3}$$

where τ_s is the surface stress, k_s is a linear drag coefficient with velocity units, and C_D is a dimensionless quadratic drag coefficient. The choice of the stress law will be discussed in more detail below. The vertical momentum flux betwen the boundary layer and the layer above it is

$$Z_{01} = M_0 w^+ - M_1 w^- + \nu(M_0 - M_1) \ , \tag{2.4}$$

where $w = w^+ - w^-$, w^+ is the nonnegative vertical velocity out of the boundary layer, w^- is the nonnegative vertical velocity into the boundary layer from the layer above it with angular momentum M_1, and ν is a vertical friction coefficient. We will assume that the only processes leading to vertical velocities at the top of the boundary layer are Ekman pumping and suction (pumping with $w < 0$), so only one of w^+ and w^- are nonzero at any given radius.

With some manipulation of equations (2.1) and (2.2):

$$\psi_0 = r\left[\frac{h_0\frac{\partial v_0}{\partial t} + k_s v_0 + C_D(u_0^2 + v_0^2)^{1/2}v_0 + (w^- + \nu)(v_1 - v_0)}{f + \frac{\partial rv_0}{r\partial r}}\right]. \qquad (2.5)$$

As was shown in Part I, if the densities of the boundary layer and the layer above it are the same, the angular momenta of these layers are the same and the terms proportional to $(v_1 - v_0)$ vanish. Even without that assumption, the vertical diffusion of angular momentum above the boundary layer will be assumed small and ignored, and further we will be primarily examining the regions of the boundary layer with positive Ekman pumping where $w^- = 0$. If the relative vorticity is negative, its magnitude is assumed smaller than the Coriolis parameter, so the total vorticity, $f + \frac{\partial rv_0}{r\partial r}$, is positive everywhere. The scaling of the system is quite different in a region with vanishingly small total vorticity and will not be considered here. Further, negative total vorticity leads to inertial instability and is not likely to last long in an azimuthal average.

We will also assume, as is the case in a tropical cyclone, that $v_0 \gg u_0$. This allows $(u_0^2 + v_0^2)^{1/2}$ to be replaced by $|v_0|$ and leads to a simple quadratic stress law. However, if in the early stages of development there are turbulent horizontal velocities in the boundary layer greater than, or of order of, the organized azimuthal flow, this will have the same effect as having a linear term in the stress. This assumes that there is some isotropic turbulence with $(u_0^2 + v_0^2)$ approximately constant. The linear stress term is a return to simple eddy viscosity theory and has already been included under this reasoning.

In calculating the vertical velocity at the top of the boundary layer, without allowing changes in the vertical thermodynamic structure, there is little difference between velocity out of the layer or an increase of the thickness of the layer. Therefore we will ignore the time derivate of the thickness in the continuity equation when calculating the Ekman pumping. Combining the equation for continuity in the

boundary layer (2.1), with a radial derivative of the equation for the lateral mass flux (2.5) yields an expression for the vertical pumping out of the boundary layer:

$$w = \frac{\partial}{r\,\partial r} \left[\frac{k_s r v_0 + C_D r |v_0| v_0 + h_0 \frac{\partial r v_0}{\partial t}}{f + \frac{\partial r v_0}{r\,\partial r}} \right] . \tag{2.6}$$

The terms due to friction (those proportional to k_s or C_D) are the Ekman pumping, though occasionally pumping from the time dependent term is included under this rubric. During periods of exponential intensification, proportional to $e^{\gamma t}$, the time derivative term has the same spatial dependence as the linear drag term. For a boundary layer a kilometer deep, a disturbance would need to have an e-folding time of about a day for the time dependent term to be comparable to the linear drag term. Since examination of the linear term provides the same information, this term will not be considered further.

If we Taylor expand eq. (2.6) with an amplitude expansion in v_0 as was used in Part II (Handel 1991b), and is likely to be used in any similar weakly nonlinear analysis, the first two terms are

$$w^{(1)} = \frac{k_s}{f} \frac{\partial r v_0}{r\,\partial r} \tag{2.7}$$

$$w^{(2)} = \frac{\partial}{f r\,\partial r} \left(C_D r |v_0| v_0 \right) - \frac{k_s}{f^2} v_0 \frac{\partial}{\partial r} \left(\frac{\partial r v_0}{r\,\partial r} \right) - k_s \left(\frac{\partial r v_0}{f r\,\partial r} \right)^2 . \tag{2.8}$$

For the lowest order stationary solutions, which are linear, the stress is then linearly proportional to the tangential velocity, and the Ekman pumping is proportional to the vorticity. This familiar result was first obtained by Charney and Eliassen (1949).

We will now examine radial profiles for the Ekman pumping of a vortex that fits the solutions found in Part I. The vortex solution is for a system with lateral viscosity that approaches solid body rotation at the central axis. Curves for the tangential velocity of the isolated vortex and the expansions for the Ekman pumping given in eqs. (2.7) and (2.8), are shown in figure 1. The solutions are nondimensionalized as in Part II, with a length scale of $L = 158\,\mathrm{km}$, a time scale of $1/f = 2 \cdot 10^4\,\mathrm{sec}$, and

a velocity scale of $Lf = 7.9\,\text{m/sec}$, while $C_D = 1 \cdot 10^{-3}$ and $k_s = C_D \cdot 5\,\text{m/sec}$. The test disturbance has unit amplitude, i.e. a maximum velocity of $7.9\,\text{m/sec}$.

The solution for $w^{(1)}$ has maximum Ekman pumping in the center accompanying the vorticity maximum. The second order terms for pumping out of the boundary layer, $w^{(2)}$, result in a radial profile with a maximum away from the center and negative values at the center. The spatial forms of each of these profiles are not dependent of the disturbance amplitude. Despite the negative values at second order, the sum of the first and second order terms remains positive in the center as long as the expansion is valid with a nondimensional amplitude less than about 0.3. An examination of the full expression for w shows that the total value for w cannot go negative at the center of a cyclonic vortex and that the negative first negative term at second order proportional to k_s only serves to reduce the linear term.

Outside of the realm of formal amplitude expansions, it is more useful to compare profiles of Ekman pumping with simple linear and quadratic stress laws. These are shown in figure 2. The Rossby number has been assumed small, so only the planetary vorticity is included in the denominator of eq. (2.6) for the calculations, and the forms of the profiles are independent of disturbance amplitude. Whether or not one includes a linear stress term leads to radically different radial profiles of the forced vertical velocity. *The linear term alone produces a maximum in the top of the boundary layer vertical velocity at the center of the vortex. For stress proportional to the square of the velocity with small Rossby number, the Ekman pumping vanishes at the center.* At large amplitude, where the quadratic stress term dominates the Ekman pumping, it is easy to see that such a radial distribution of forced vertical velocity would lead to a well defined eyewall, as observed, if positive Ekman pumping serves to initiate deep convection.

For very strong disturbances, some workers (e.g. Moss and Rosenthal 1975) have suggested a drag law that increases faster than quadratically, with $C_D = C_{D0}(1+av_0)$.

This adds a cubic drag term. The radial profile of Ekman pumping for a cubic drag term alone, shown in figure 2, is quite similar to that for a quadratic term alone. The locations of the zero crossings, maxima, and minima; and ratios of maxima to minima are within 15%. So, inclusion of a C_D with linear velocity dependence enhances the Ekman pumping, especially for large velocities, but does little to the spatial structure.

We now turn briefly to the slab symmetric case. Results for this case can be easily generated by changing all derivatives of the form $\frac{\partial rv}{r \partial r}$ to $\frac{\partial v}{\partial x}$ in eq. (2.6) and making all dependent variables functions of x rather than r. Using the same spatial profile for the velocity profile as we did for the cylindrical case (though this profile is no longer a solution to a system of equations relating to cyclones), figure 2 provides a plot of the Ekman pumping for small Rossby number for a linear drag law. The greatest difference between this and the axisymmetric case is the weaker pumping in the high vorticity region due to the lack of the large curvature term near the vortex center. On the other hand, Ekman suction away from the center is greater for the slab symmetric case. For higher order drag laws, the differences are less dramatic.

Over the extent of a disturbance domain, with the velocity vanishing at the boundaries, it is clear that the integral of the Ekman pumping must vanish as long as the assumption that the total vorticity remains positive holds. For the axisymmetric case, the area with Ekman suction is very large (growing as r^2), so the magnitude of the suction is small. Comparatively for the slab symmetric case, the suction covers a lesser area and is several times greater.

The numerical results presented above ignore the relative vorticity in the denominator of eq. (2.6). This has the advantage of providing profiles that are amplitude independent. However, the profiles for Rossby number not much less than one are, of course, dependent on disturbance amplitude. Figure 3 shows a sequence of curves of Ekman pumping for increasing vortex intensities. All of the nonlinear terms of

eq. (2.6) are included. The weakest vortex has Ekman pumping not greatly different from the linear case. At a strength of 0.5, the maximum pumping has moved far from the center. For strong vortices, there is positive pumping in region of negative vorticity. For the case of vortex strength of 4.0 the maximum postive pumping is in a region of negative vorticity. This is due in large part to the behavior of the denominator of of the argument in eq. (2.6) in regions where the vorticity gradient is strongly negative. Vortices of greater strength would violate the assumption of positive total vorticity in some regions. However, in actual tropical cyclones the magnitude of the vorticity gradient in the regions of negative relative vorticity is rarely this large.

3.   Discussion and Conclusions

In the course of this analysis we assumed that the hurricane prone regions of the tropics in summertime are conditionally unstable and that positive Ekman pumping initiates deep convection. We further assumed that the primary effect of organized convection is entrainment and the import of angular momentum toward the convecting region. (See Part I for discussion of these assumptions.) Such import leads to the large observed tangential winds. This in turn leads to a change in the thermal structure through geostrophic adjustment by the radiation of gravity waves. Probably both the radial profile of Ekman pumping and upper level stabilization are responsible for the lack of convection in the center of a tropical cyclone. However, only control by Ekman pumping is consistent with the observation that the eyewall region is narrow and well defined.

In a weak vortex, the presence of turbulent velocities in the boundary layer greater than the large scale organized velocity leads to an effective linear drag at the surface. This in turn implies Ekman pumping proportional to vorticity. For a

weak, viscous vortex, approaching solid body rotation near the center, the vorticity maximum is at the vortex axis, so the convection is also a maximum near the center. In a more intense vortex, the organized velocity dominates and the drag felt by the large scale disturbance becomes more closely proportional to the square of the organized velocity. Under these conditions, the Ekman pumping vanishes at the center and reaches a maximum at some finite radius producing the characteristic eye of the hurricane.

Observational results of the relationship of low level vorticity and deep convection are mixed. In analysis of GATE (GARP Atlantic Tropical Experiment) data, Reeves *et al.* (1979) found good correlation between 700 mb vorticity and precipitation in the better organized systems of Phases II and III, but not in Phase I. Chen and Ogura (1982) claimed little correlation using the same basic data set and therefore argued against CISK. However, 700 mb is not the appropriate level for measuring vorticity to reach this conclusion. In composites of developing disturbances, Williams and Gray (1973) found that a low level vorticity maximum was at the center of tropical cloud clusters. Later workers at Colorado State University (e.g. McBride and Zehr 1981) found cloud clusters were associated with relative vorticity maxima at 900 mb on scales of about 200 km. In view of this inconsistent picture, there is need for continued study on the relationship between deep tropical convection and large scale flow.

For intense cyclones, there are countless observations showing the annular structure of the eyewall; this is clearest in radar scans (e.g. Franklin *et al.* 1988). Many detailed observations have shown that the eyewall is a region of mean ascent and boundary layer convergence (e.g. Jorgensen 1984). Once the quadratic stress term dominates, sharper gradients in Ekman pumping develop and a narrower convective region develops. This forces sharper gradients of velocity in the vortex. Continued spatial collapse is probably only limited by internal momentum diffusion processes.

Though we concentrated on cylindrical distrubances, weak tropical disturbances often have much of their vorticity in shear rather than curvature (Williams and Gray 1973). As a disturbance evolves, the development of curvature in easterly waves encourages development by causing greater Ekman pumping for the same vortex strength. This is made clear by comparison of the Ekman pumping in the extreme cases of cylindrical and slab symmetry using the same disturbance velocity profile. The difference reaches a factor of two greater for the cylindrical case at the center for the updraft. Further, shorter wavelength disturbances of the same amplitude are more likely to develop since these have greater Ekman pumping for the same strength disturbance.

Most theoretical analyses are performed on linear or weakly nonlinear systems for convenience of calculation. On the other hand, numerical models almost all use quadratic, or quadratic plus cubic, drag laws. Different choices of surface drag lead to radical changes in the forced vertical motion. Though we have argued that for weak circulation, a linear drag law may in fact be appropriate, is is probably unwise to rely on ease of calculation for use of linear drag laws in theoretical calculations without further justification. In axisymmetric or slab symmetric models, it may be unwise to use only a quadratic drag law in numerical models, unless examining only very strong disturbances. This is less a problem in three dimensional models with high resolution, since small scale motions provide some of the same behavior modeled here with a linear drag law.

For vertically stable conditions, where Ekman pumping provides a weak decay mechanism for the large scale dynamics, the spatial details of Ekman pumping may not be of great significance. However, for regions with a conditionally unstable boundary layer, where Ekman pumping controls the location of deep precipitating convection and the location of most of the convergence aloft, the spatial dependence of the pumping strongly effects the further development of a disturbance, so stress

boundary conditions must be chosen with care.

Acknowledgments

References

Abdullah, Abdul Jabbar 1953: *On The Dynamics of Hurricanes.* I. The eye of the hurricane; *Meteorological Papers* **2** (2), New York Univ. Press, New York, 7–24.

Anthes, Richard A. 1982: *Tropical Cyclones — Their Evolution, Structure and Effects; Meteorological Monographs* **19** (41), Americal Meteorological Society, Boston, Massachusetts, 208pp.

Byers, H. R. 1944: *General Meteorology,* McGraw-Hill, New York.

Charney, J. G. and A. Eliassen 1949: A numerical method for predicting the perturbation of the middle latitude westerlies; *Tellus* **1** (2), 38–54.

Chen, Yi-Leng and Yoshi Ogura 1982: Modulation of convective activity by large-scale flow patterns observed in GATE; *J. Atmos. Sci.* **39** (6), 1260–1279.

Dampier, William 1699: *Voyages and Descriptions;* vol. 2 of *A New Voyage round the World,* Part I. Supplement of the Voyage round the World, Knapton, London.

Eliassen, A. 1971: On the Ekman layer in a circular vortex; *J. Meteor. Soc. Japan* (Ser. II) **49** (Special issue), 784–789.

Eliassen, Arnt and Magne Lystad 1977: The Ekman layer of a circular vortex. A numerical and theoretical study; *Geophysica Norvegica* **31** (7), 1–16.

Emanuel, Kerry A. 1986: An air-sea interaction theory for tropical cyclones. Part I: Steady-state maintenance; *J. Atmos. Sci.* **43** (6), 585–604.

Franklin, James L., Stephen J. Lord, and Frank D. Marks, Jr. 1988: Dropwindsonde and radar observations of the eye of Hurricane Gloria (1985), *Mon. Wea. Rev.* **116** (5), 1237-1244.

Greenspan, H. P. and L. N. Howard 1963: On a time-dependent motion of a rotating fluid; *J. Fluid Mech.* **17** (3), 385-404.

Gyakum, John R. 1983: On the evolution of the QE II Storm. I: Synoptic aspects; *Mon. Wea. Rev.* **111** (6), 1137-1155.

Handel, Mark David 1991a: Tropical cyclone intensification. Part I: Impediment to linear growth, ms, Phillips Laboratory /GPAS, Hanscom AFB, Massachusetts, December, to be submitted to *J. Atmos. Sci.*

Handel, Mark David 1991b: Tropical cyclone intensification. Part II: Growth from finite amplitude disturbances, ms, Phillips Laboratory /GPAS, Hanscom AFB, Massachusetts, December, to be submitted to *J. Atmos. Sci.*

Hawkins, Harry F. and Daryl T. Rubsam 1968: Hurricane Hilda, 1964. II. Structure and budgets of the hurricane on October 1, 1964, *Mon. Wea. Rev.* **96** (9), 617-636.

Jorgensen, David P. 1984: Mesoscale and convective-scale characteristics of mature hurricanes. Part II: Inner core structure of Hurricane Allen (1980); *J. Atmos. Sci.* **41** (8), 1287-1311.

Kuo, H. L. 1959: Dynamics of convective vortices and eye formation; in Bert Bolin, ed., *The Atmosphere and the Sea in Motion* (also referred to as *The Rossby Memorial Volume*), Rockefeller Inst. Press with Oxford Univ. Press, New York, 413-424.

Kurihara, Yoshio and Morris A. Bender 1982: Structure and aanalysis [sic] of the eye of a numerically simulated tropical cyclone; *J. Meteor. Soc. Japan* **60** (1), 381–395.

Malkus, Joanne S. 1958: On the structure and maintenance of the mature hurricane eye; *J. Meteor.* **15** (4), 337–349.

Moss, Michael S. and Stanley L. Rosenthal 1975: On the estimation of planetary boundary layer variables in mature hurricanes; *Mon. Wea. Rev.* **103** (11), 980–988.

Ooyama, Katsuyuki 1969: Numerical simulation of the life cycle of tropical cyclones; *J. Atmos. Sci.* **26** (1), 3–40.

Palmén, E. 1948: On the formation and structure of tropical hurricanes; *Geophysica* **3**, 26–38.

Rasmussen, E. 1979: The polar low as an extratropical CISK disturbance; *Quart. J. Roy. Meteor. Soc.* **105** (445), 531–549.

Reeves, Robert W., Chester F. Ropelewski, and Michael D. Hudlow 1979: Relationships between large-scale motion and convective precipitation during GATE; *Mon. Wea. Rev.* **107** (9), 1154–1168.

Schubert, Wayne H., James J. Hack, Pedro L. Silva Dias, and Scott R. Fulton 1980: Geostrophic adjustment in an axisymmetric vortex; *J. Atmos. Sci.* **37** (7), 1464–1484.

Smith, Roger K. 1980: Tropical cylone eye dynamics; *J. Atmos. Sci.* **37** (6), 1227–1232.

Syōno, Sigekata 1951: On the structure of atmospheric vortices; *J. Meteor.* **8** (2), 103–110.

Williams, Knox T. and William M. Gray 1973: Statistical analysis of satellite-observed trade wind cloud clusters in the western North Pacific; *Tellus* 25 (4), 313–336.

Willoughby, H. E. 1979: Forced secondary circulations in hurricanes; *J. Geophys. Res.* 84 (C6), 3173–3183.

Yamasaki, Masanori 1977: The role of surface friction in tropical cyclones; *J. Meteor. Soc. Japan* 55 (5), 559–571.

Figure 1: Ekman pumping for weakly nonlinear expansions. Curve (b) shows the Ekman pumping at first order, $w^{(1)}$; while curve (c) shows the second order correction, $w^{(2)}$. Curve (a) shows the tangential velocity profile of the vortex used divided by 1000. The vertical dashed lines are at the radius of maximum winds and the radius of zero relative vorticity. For each curve the vortex is assumed to have unit nondimensional amplitude corresponding to a maximum dimensional velocity of 8 m/sec and the length unit is 160 km.

Figure 2: Comparison of Ekman pumping between different drag laws and comparison between cylindrical and slab symmetries. Curve (a) shows the Ekman pumping for a linear drag law in cylindrical symmetry [same as (b) in previous figure]; curve (b) shows the Ekman pumping a quadratic drag law in cylindrical symmetry; curve (c) shows the Ekman pumping for a cubic drag law in cylindrical symmetry; and curve (d) shows the Ekman pumping for a linear drag law in slab symmetry. All calculations include only the planetary vorticity in the denominator. Scaling is as in figure 1.

Figure 3: Ekman pumping for various vortex strengths. Scaling is as in figure 1. N.B. the difference in vertical scale between the plots.
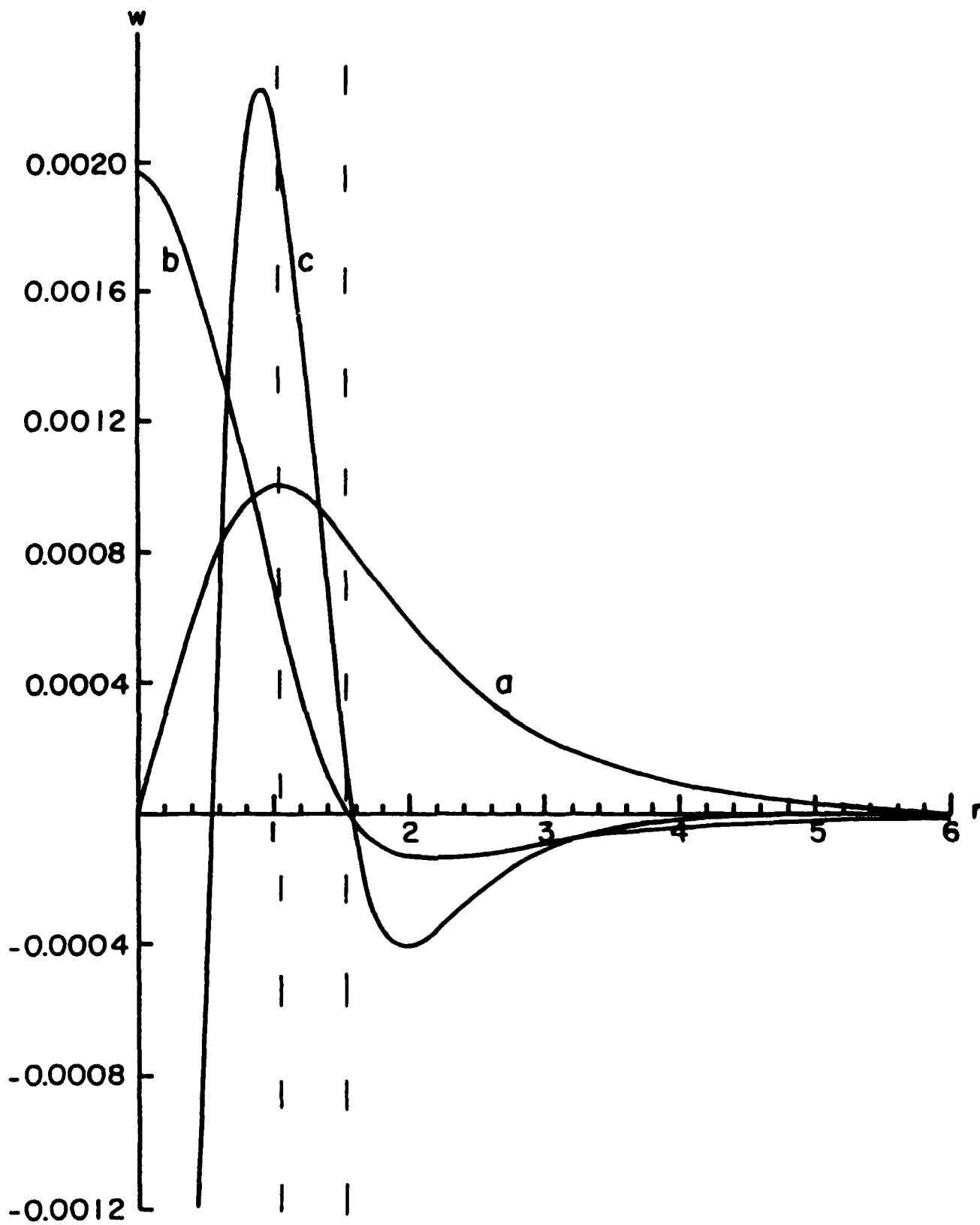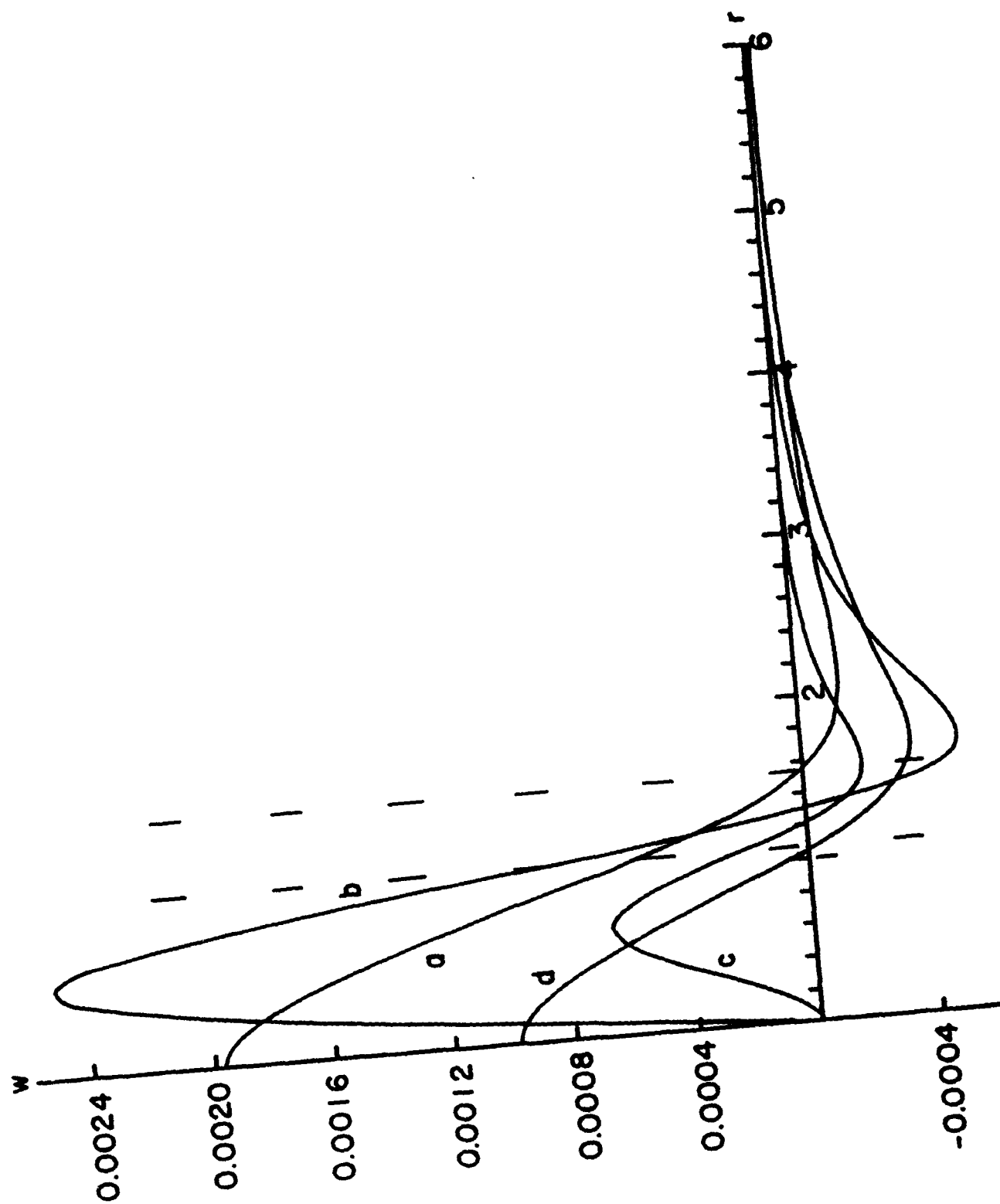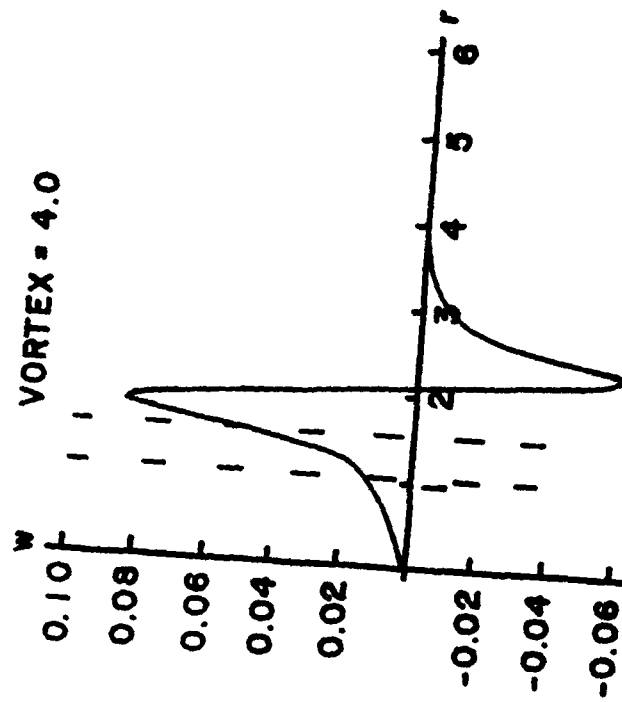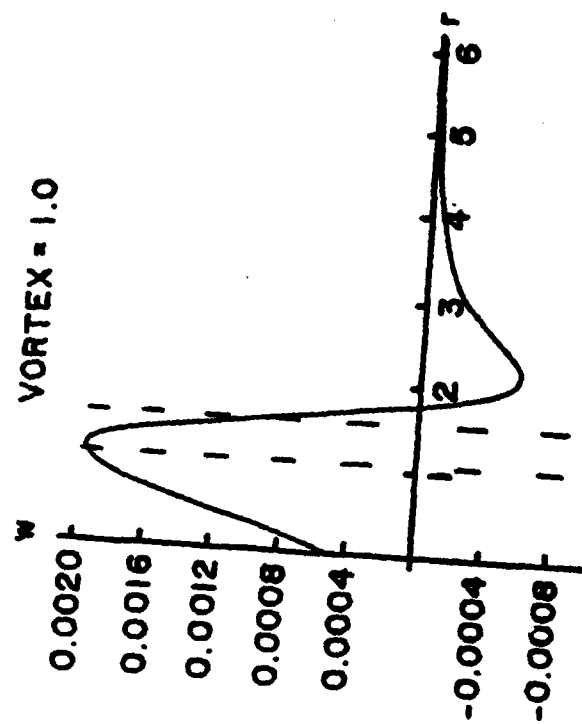
FIGURE 1

FIGURE 2

VORTEX = 0.1
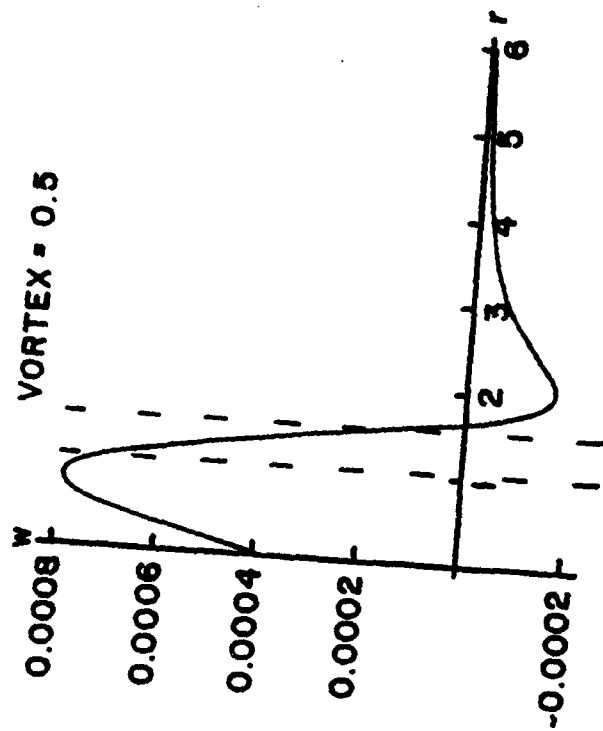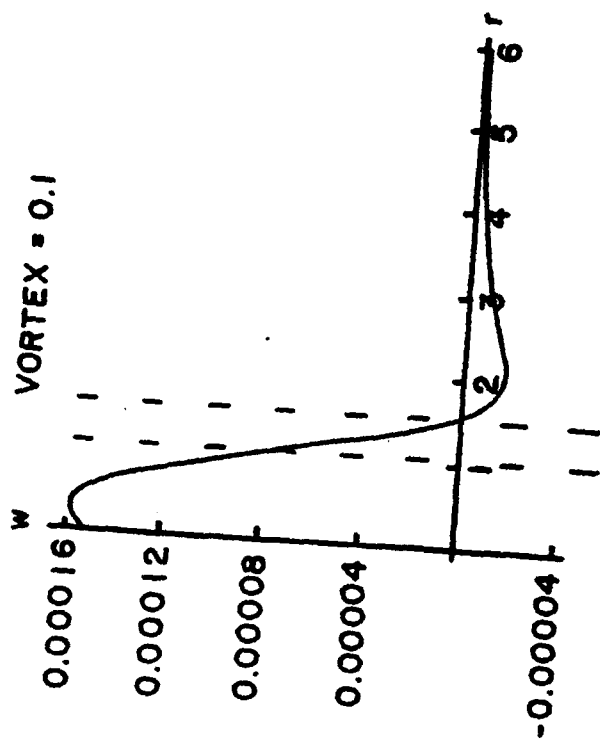VORTEX = 0.5
VORTEX = 1.0
VORTEX = 4.0

FIGURE 3

135

1991–1992 SCEEE Geophysics Scholar Program

Sponsored by

Department of the Air Force
Phillips Laboratory / Geophysics Directorate

Conducted by the

Southeastern Center for
Electrical Engineering Education

Final Report

Part II. On the Energetics of
Asymmetric Tropical Cyclone Outflows

Prepared by:	Dr. Mark David Handel
Research Location:	Phillips Laboratory / Geophysics Directorate
Phillips Laboratory Research Contact:	Dr. J. William Snow
Date:	May 29, 1992
Contract Number:	F19628-86-C-0224

Abstract

Tropical cylone outflows have been known for some time to be asymmetric or even jet-like. On the other hand, the low level inflows are very close to cylindrical symmetry. Some workers have argued that the asymmetry aloft is induced through interactions with the larger scale environmental flow. However, we believe that tropical cyclones naturally de lop strong asymmetries in the outflow. Simple energy arguments are presented to show that narrow outflow regions are energetically favored states; i.e. states that require the least amount of kinetic energy for a given convectively driven mass flux. The very simple model used assumes that angular momentum is conserved and that the organized convection of a cyclone eyewall draws mass in at low levels and expels it aloft. Over a large range of mass fluxes (and hence, storm lifetimes), a robust result is obtained that the outflow regions should cover arcs of between 30 and 40 percent of a circle and that the inflow should cover the entire circle (i.e. be symmetric). This is consistent with observed outflows.

1   Introduction

Once a tropical cyclone reaches moderate intensity, in the lower levels there is inflow from all directions toward the center. However, in the upper troposphere where there is mean outflow from the center, the outflow is highly asymmetric (deviations from cylindrical symmetry) and is usually concentrated into one or two sectors covering less than half the circle. Many works have attributed the outflow asymmetries and the associated angular momentum eddy fluxes to interactions between the tropical cyclone and the surrounding synoptic scale flow. We show here that asymmetric outflows are energetically favored for an isolated tropical cyclone. This suggests that the asymmetry is likely to be intrinsic to the tropical cyclone phenomenon.

Ferrel (1856) was probably the first to realize that inflow towards the center of a cyclone would result in spiral pathways for parcels and would result in spin-up of a storm. He also realized that in an intense cyclone the outflow would be a weakening spiral, but failed to realize that anticyclonic motion at greater radius would also result. Until the work of Haurwitz in 1935, it was generally thought that the circulation of tropical cyclones extended upwards only three kilometers, or less. Haurwitz showed that for the observed central pressure perturbations, reasonable limits on the lapse rate implied that the cyclones reached at least ten kilometers, if not higher. He also noted that a storm's circulation might continue above the level of zero pressure gradient into a level of negative pressure gradient and outflow. Durst and Sutcliffe (1938) also suggested that

there might be an anticyclonic circulation during the early stages of development due to a pressure increase from the accumulation of mass in the core region. Through they developed equations showing the consequenses of conservation of angualr momentum near the surface, they failed to apply the same logic to the outflow level.

E. S. Jordan (1952) provided perhaps the first conclusive observations of anticyclonic flow aloft outside of the core region, though this was previously hinted at by Riehl (1951). Jordan also showed that the outflow was not symmetric, though the use of composited data left some room for doubt. She found a single outflow region in Earth bound coordinates, but two outflow regions in storm moving coordinates. These observations indicated that at small radius the upper level tangential flow was cyclonic, while at larger radius the flow became anticyclonic. Descriptions of outflows based on satellite derived winds and aerological data from five different storms were presented by Black and Anthes (1971). All of the storms they studied had highly asymmetric distributions in the radial component of the wind in the upper layer.

Three dimensional numerical models also develop asymmetries. The simulation of Anthes (1972) developed strong asymmetries, even with initial conditions of an axisymmetric vortex in gradient wind balance broken only by roundoff errors and the imposition of a square domain. Similar results were obtained by Kurihara and Tuleya (1974) and by Jones (1977), who also started with initial conditions of axisymmetric vortices. In all of the simulations, eventually wave number two dominated the asymmetric components, though at early times higher wavenumbers were the most energetic. Anthes' (1971) simulations had initial asymmetries at wave number 4, corresponding to symmetry breaking by the cylindrical cyclone in a square domain. However, before long the asymmetries were dominated by wave numbers 1 and 2. However, these modelled asymmetries were usually not as strong as observed ones (Merrill 1988a).

The outflow asymmetries are often attributed to external influences or from storm

motion (e.g. Miller 1958). There have been several recent works that have taken the view that the asymmetries in tropical cyclone outflows are to a large part environmentally induced (e.g. Molinari and Vollaro 1989, 1990). Further, there have been claims that the momentum surges from these external influences play a significant role in the intensification of cyclones. The simulations of Challa and Pfeffer (1980, 1984, 1990) have shown rather conclusively that the eddy angular momentum fluxes aloft encourage intensification and are even essential for weak disturbances to exceed the threshold for intensfication (1990). Though most of the works examining momentum fluxes from external influences have concluded that these fluxes assist the cyclone intensification process, Merrill (1988b) concludes that the sum of all the external influences (including vertical shear) must be a negative influence since so few storms reach their thermodynamically limited maximum intensity.

In the simple calculations presented below, we will examine the energy of the outflow region of a tropical cyclone for various configurations. The vertical component of the angular momentum (i.e. the angular momentum due to the tangential circulation) will be assumed absolutely conserved for each parcel. The sole purpose of the eyewall will be to draw air in at the lower level and expel it aloft.

2 Energetics of Spin-up

The very simple model used here has only two layers, a lower layer with mean inflow and an upper layer with mean outflow. We will assume that all fluid is initially at rest in a rotating system with constant Coriolis parameter, f. The coordinate system is cylindrical, with radius r. Only azimuthal velocities, v, within this coordinate system will be considered. The vertical component of the angular momentum, M, of a parcel is then:

$$M \equiv rv + \frac{f}{2}r^2 .$$

(1)

In the initial state parcels are at rest with respect to the earth, so a parcel at initial radius r_i has

$$M_i = \frac{f}{2} r_i^2 . \tag{2}$$

For parcels conserving their angular momentum, their azimuthal velocity at a later radial position, r_f, is

$$v(r_f) = \frac{f}{2r_f}(r_i^2 - r_f^2) . \tag{3}$$

Changes in the altitude or density of a parcel do not alter this.

There will be no examination of dynamics here. We now assume a process that sucks in fluid in the lower layer of the system at radius r_0 and ejects it aloft at the same radius. (Any similarity between this process and cumulus convection in an eyewall is intentional, but limited.) The process may selectively pull fluid from only a limited fraction of the sectors and similarly may eject fluid into only limited sectors. The geometry of the process is illustrated in figure 1. At the conclusion of the process, fluid will have been sucked in from sectors covering a fraction ϵ_1 of the annulus from r_0 to r_1. The same fluid will have been ejected into sectors covering fraction ϵ_2 of the annulus between r_0 and r_2. The choices of the number and symmetry of sectors are arbitrary and will have no effect on the following calculation. Transport of the fluid will be assumed to occur in "conveyor belt" fashion. Hence parcels that begin at radius r_0 in the lower layer finish at radius r_2 aloft, and parcels that begin at radius r_1 in the lower layer finish at radius r_0 aloft. For parcels of arbitrary initial radius in the interval $[r_0, r_1]$, the initial and final positions satisfy

$$\epsilon_1 \rho_1 h_1 \pi (r_i^2 - r_0^2) = \epsilon_2 \rho_2 h_2 \pi (r_2^2 - r_f^2) , \tag{4}$$

141

where ρ_j and h_j are the density and depth of layer j, respectively; and the lower layer is designated by $j = 1$, while the upper layer has $j = 2$. For each level the total masses are the same:

$$V = \epsilon_1 \rho_1 h_1 \pi (r_1^2 - r_0^2) = \epsilon_2 \rho_2 h_2 \pi (r_2^2 - r_0^2) . \tag{5}$$

One can now easily solve for the tangential velocities in the outflow:

$$v_2(r_f) = \frac{f}{2r_f} \left[r_1^2 + \gamma r_0^2 - (1 + \gamma) r_f^2 \right] , \tag{6}$$

where

$$\gamma \equiv \frac{\epsilon_2 \rho_2 h_2}{\epsilon_1 \rho_1 h_1} . \tag{7}$$

The pressure thicknesses of the inflow and outflow layers are observed to be about equal at approximately 200 mb, with most of the inflow concentrated in a lower layer roughly from 900–700 mb and the upper layer outflow roughly confined within 300–100 mb. Therefore $\rho_1 h_1 \approx \rho_2 h_2$ and the nondimensional ratio $\gamma \approx \epsilon_2/\epsilon_1$. An example velocity profile is shown in figure 2.

From the upper layer velocity profile one can obtain the relative vorticity:

$$\zeta_2 = \frac{\partial r_f v_2}{r_f \, \partial r_f} = -f(1 + \gamma) . \tag{8}$$

In the outflow the vorticity is independent of radius and strongly negative. Even the absolute vorticity is negative and equal to $-f\gamma$. Hence, the smaller the ratio γ the less negative the vorticity.

The *deus ex machina* process described above, though it conserves angular momentum, does not conserve energy. In fact large amounts of energy are needed to drive the postulated mass flux. The remainder of this calculation is concerned with determining the required energy and the dependence of the energy on the mass flux, inflow fraction, and outflow fraction.

We will not be concerned here with fluid initially at rest in the upper level moving outwards or fluid in the lower level moving only inwards. In both of these cases the motion is all strictly away from the radius of zero tangential velocity and so symmetric motion clearly requires the least energy. Also, the return branch of fluid sinking without much change to its radial position is ignored here.

In the outflow layer aloft, the integrated kinetic energy is

$$K_a = \frac{1}{2} \int_{r_0}^{r_1} \epsilon_2 \rho_2 h_2 v_2^2(r_f) \, \pi r_f \, dr_f \tag{9}$$

for fluid that was initally at rest in the lower level. Since we are interested in the difference in energy for differenct configurations of outflow for a given amout of mass flux, it is easier to examine the kinetic energy density. Substitution by eqs. (3)–(7), followed by an integration, and a fair amount of algebraic manipulation yields for the kinetic energy density:

$$T_a \equiv \frac{K_a}{V} = \left(\frac{fr_0}{4}\right)^2 \left[\left(\frac{1+\gamma}{2}\right)\left(\frac{B}{\gamma} - 3B - 2[1+\gamma]\right)\right.$$
$$\left. + \frac{\gamma}{B}(B+1+\gamma)^2 \ln\left(1 + \frac{B}{\gamma}\right)\right], \tag{10}$$

where the nondimensional parameter

$$B \equiv \frac{r_1^2 - r_0^2}{r_0^2} \propto \frac{V}{\epsilon_1}. \tag{11}$$

Some limits are of note:

$$\lim_{\gamma \to 0} T_a = +\infty, \text{ for } B > 0$$

$$\lim_{\gamma \to \infty} T_a = +\infty, \text{ for } B > 0$$

$$\lim_{B \to 0} T_a = 0, \text{ for } \gamma > 0$$

$$\lim_{B \to \infty} T_a = +\infty, \text{ for } \gamma > 0.$$

These limits can be obtained through a combination of direct substitution and application of l'Hopital's rule.

A contour plot of T_a, as a function of γ and B is provided in figure 3. The leading dimensional factor in the expression for T_a has been omitted, so the units are arbitrary. Important to note is that the figure is independent of r_0 (eyewall radius) and f (latitude). Though both ϵ_1 and ϵ_2 are each in the interval $(0, 1]$, there is no such restriction on the magnitude of γ.

We now make the *ansatz* that the most likely states are those of lowest energy. It is visually apparent that the mean energy *density* increases about linearly in B. Paths in the space of $B \times \gamma$ that vary the inflow fraction ϵ_1 for constant mass flux, V, and constant outflow fraction, ϵ_2, are straight lines emanating from the origin. For any given amount of fluid moved by our cylindrical pump, the energy density K_a is minimized along these paths by heading towards the origin. These paths can only be followed until the inflow sector fraction, ϵ_1, reaches its maximum value of 1. In this limit then $\gamma = \frac{\epsilon_2 \rho_2 h_2}{\rho_1 h_1} \approx \epsilon_2$, and we see the optimal sector fraction for the outflow aloft is about 0.4, corresponding to a total angle of about 150°, for young disturbances, decreasing to about 0.3 as the amount of fluid pumped increases. This result is fairly insensitive to the amount of fluid pumped, though the selection becomes more strongly favored by this energetic criterion as the system develops.

In the limit of large B, the preferred outflow fraction ratio, γ, approximately satisfies the implicit equation

$$2 \ln \left(\frac{B}{\gamma} \right) = 5 + \frac{1}{\gamma^2} \, . \tag{12}$$

This was determined by setting $\frac{\partial T_a}{\partial \gamma} = 0$, assuming $B \gg 1$, and keeping only the highest order terms in B. A curve of the solution to eq. (12) is shown in figure 4 along with a curve for the preferred γ as determined by the full equation (11) and figure 3. For all but the smallest values of B, eq. (12) provides an excellent approximation.

3 Discussion and Conclusions

Though it has often been suggested that a strong anticyclone aloft encourages development of a tropical cyclone (e.g. Merrill 1988b), we take a contrary position. The creation of any anticyclonic flow is a consequence of the outflow driven by the intensification process and is an energetic drain on the system. (Claims that the outflow circulation and the anticylone aloft can serve as a source of intensification inspired the inclusion of the leading epigraph.) Though surface concentrated perturbations can drive positive feedbacks through air/sea interactions (Emanuel 1986, Handel 1991), the upper level flow provides no energy sources to drive intensification. It is because there are no positive feedbacks between a strong flow aloft and further tropical cyclone intensification that the energy minimization heuristic used here has any applicability.

The inital radius of a parcel in this problem is also what is referred to as the *potential radius* in some works that make use of angular momentum coordinates. It is the location where a parcel has no kinetic energy in the rotating reference frame. To a crude approximation, the magnitude of the velocity increases linearly with deviation from this postion. Hence, kinetic energy increases about quadratically with change in radial position. Since the radial ordering of the parcels is fixed, maintained strictly in the problem here though slightly less well in nature, it is impossible for all parcels to return to their potential radius after being lifted aloft and their radial ordering reversed. Outflow in narrow regions allows more parcels to get closer to their potential radius and therefore results in less kinetic energy aloft.

Narrow outflows have another profound effect on the kinetic energy of the upper level anticyclone. These configurations lead to much larger internal gradients of velocity and greatly increase the lateral momentum mixing. This mixing also has the effect of reducing the strengh of the anticyclone as well as reducing the strength of the cyclonic

circulation at smaller radius aloft. Such mixing encourages the intensification of a tropical cyclone (Handel 1991). Though in the calculation above angular momentum was strictly conserved by each parcel, a more complete calculation that allowed for more momentum mixing by states of greater asymmetry would show even stronger preference for states with narrow outflows.

The asymmetries are so great that Merrill (1988a) notes that for some storms there are regions of large scale *inflow* in what is referred to as the "outflow layer", and that in some storms these can contribute as much or more eddy import of angular momentum than is found in the outflow regions. As can be seen in budgets of composites he provided, the tendency of the mean ouflow to export planetary angular momentum (in a cyclone reference frame), is partially countered by both the mean transport of relative angular momentum and eddy transports of angular momentum. At large radius (greater than 1000 km), it appears that the eddy transport is larger than the mean.

The notion that large eddy angular momentum fluxes imply strong external influences on tropical cyclones, such as that made in Molinari and Vollaro 1989, makes a large and unjustified leap. However, the observational result of Molinari and Vollaro that tropical cyclone intensification is well correlated with increases in upper level eddy angular momentum fluxes makes dynamical sense. It is consistent with this work showing that asymmetric outflows require less energy to drive as well as the results of Handel (1991) that shows that increases in the eddy diffusivity in the upper layer of a simple tropical cyclone model encourages intensfication. However, since asymmetric states are energetically favored by processes inherent to tropical cyclone mechanics, the eddy fluxes of angular momentum are likely due to *internal* cyclone processes.

The energy based arguments for asymmetric outflow provide no dynamical method of creating these asymetries. Various instabilities have been proposed for the upper levels of tropical cyclones (e.g. Alaka 1961, Anthes 1972). Though the velocity profile

from "conveyor belt" flow is inertially unstable in the outflow layer, inertial instabilities are symmetric. Anthes (1972) has specifically looked at instabilities that might lead to asymmetries, but not in any depth. Since the model used here provides specific velocity profiles, it is worth examining in detail the stability of the flows to asymmetric perturbations. This will part of the continuing line of our research.

Fast moving upper level waves do not appear to have sufficient energy to provide a strong direct impetus for tropical cyclone intensification. However, such upper level disturbances may help break the symmetry of the upper level outflow. This might make the low energy asymmetric states accessible.

Since there is so little in the way of dynamics in the model used and also since the energy minimization criteria is not a strict indication of dynamical development, the results obtained can only be considered suggestive. Nevertheless, the following conclusions are indicated. For a given mass pumping by a tropical cyclone eyewall, the least energy is required if the inflow is symmetric and the outflow covers 30–40% of the sectors of the outflow region. If the pressure thickness of the inflow and outflow layers were not approximately equal, the preferred outflow fraction is scaled by the ratio of the inflow thickness to the outflow thickness. The energetically preferred outflow fraction is robust over very large variations in mass pumping, with the preferred outflow fraction decreasing only very slowly in time after an initial rapid decrease. Further, the energetically preferred outflow fraction depends on neither cyclone latitude nor eyewall radius.

References

Alaka, M. A. 1961: The occurrence of anomalous winds and their significance; *Mon. Weather Rev.* **89** (11), 482–494.

Anthes, Richard A. 1972: Development of asymmetries in a three-dimensional numerical model of the tropical cyclone; *Mon. Weather Rev.* **100** (6), 461–476.

Black, Peter G. and Richard A. Anthes 1971: On the asymmetric structure of the tropical cyclone outflow layer; *J. Atmos. Sci.* **28** (8), 1348–1366.

Challa and Pfeffer 1980: Effects of eddy fluxes of angular momentum on model hurricane development; *J. Atmos. Sci.* **37**, 1603–1618.

Challa, Malakondayya and Richard L. Pfeffer 1984: The effect of cumulus momentum mixing on the development of a symmetric model hurricane; *J. Atmos. Sci.* **41** (8), 1312–1319

Challa, Malakondayya and Richard L. Pfeffer 1990: Formation of Atlantic hurricanes from cloud clusters and depressions; *J. Atmos. Sci.* **47** (7), 909–927.

Durst, C. S. and R. C. Sutcliffe 1938: The importance of vertical motion in the development of tropical revolving storms; *Quart. J. Roy. Meteor. Soc.* **64** (273), 75–84; with addendum on p. 240.

Emanuel, Kerry A. 1986: An air-sea interaction theory for tropical cyclones, *J. Atmos. Sci.* **43** (6), 585–604.

Ferrel, W. 1856: An essay on the winds and the currents of the ocean, *Nashville J. Medicine and Surgery* **9** (4,5). Reprinted in 1882: *Professional Papers of the Signal Service* **12**, U.S. War Dept., Washington, D.C., 7–19.

Frank, W. M. 1977: The structure and energetics of the tropical cyclone. Part II. Dynamics and Energetis, MWR 105, 1136–1150.

Handel, Mark David 1991: Tropical cyclone intensification. Part II: Growth from finite amplitude disturbances, ms, Phillips Laboratory, Hanscom AFB, Massachusetts, 54pp., submitted to *J. Atmos. Sci.*

Haurwitz, Bernhard 1935: The height of tropical cyclones and of the "eye" of the storm; *Mon. Weather Rev.* **63** (2), 45–49.

Jones, R. W. 1977: A nested grid model for a three-dimensional model of a tropical cyclone; *J. Atmos. Sci.* **34** (10), 1528–1553.

Jordan, Elizabeth S. 1952: An observational study of the upper wind-circulation around tropical storms; *J. Meteor.* **9** (5), 340–346.

Kurihara, Y. and R. E. Tuleya 1974: Structure of a tropical cyclone developed in a three-dimensional numerical simulation model; *J. Atmos. Sci.* **31** (4), 893–919.

Merrill, R. T. 1988a: Characteristics of the upper-tropospheric environmental flow around hurricanes; *J. Atmos. Sci.* **45** (11), 1665–1677.

Merrill, R. T. 1988b: Environmental influences on hurricance intensification; *J. Atmos. Sci.* **45** (11), 1678–1687.

Miller, Banner I. 1958: The three-dimensional wind structure around a tropical cyclone, National Hurricane Research Project, Report No. 15, U.S. Department of Commerce, Weather Bureau, Washington, D.C., 41pp.

Molinari, John and David Vollaro 1989: External influences on hurricane intensity. Part I: Outflow layer eddy angular momentum fluxes; *J. Atmos. Sci.* **46** (8), 1093–1105.

Molinari, John and David Vollaro 1990: External influences on hurricane intensity. Part II: Vertical structure and response of the hurricane vortex; *J. Atmos. Sci.* **47** (15), 1902–1918.

Rasmussen, Erik 1985: A case study of a polar low development over the Barents Sea; *Tellus* **37A** (5), 407–418. [near end??]

Riehl, Herbert 1951: Aerology of tropical storms; in Thomas F. Malone (ed.), *Compendium of Meteorology*, American Meteorological Society, Boston, Massachusetts, 902–913.

Figure 1: Induced mass flow from convective processes. The upper figure shows a cross section schematic of the assumed radial mass flow from the postulated covection. Flow is concentrated in a lower inflow layer and an upper outflow layer. The lower figure shows a planform view of the allowed inflow inflow (left) and outflow (right) regions. Flow is only allowed in the unshaded areas. The choices of the number and symmetry of these disconnected areas are arbitrary. The unshaded area for the inflow covers a fraction ϵ_1 of the annulus between r_0 and r_1. The unshaded area of the outflow covers a fraction ϵ_2 of the annulus between r_0 and r_2.

Figure 2: An upper level velocity profile resulting from "conveyor belt" flow. The parameters are set at $r_0 = 20\,\mathrm{km}$, $r_1 = 20r_0$, $f = 0.00002$, and $\gamma = 0.3$.

Figure 3: Normalized energy density $[T_a/(fr_0/4)^2]$ as a function of upper to lower level fraction ratio, γ, and nondimensionalized areal extent of inflow, B. Contours run from 1000 to 20000 in increments of 1000.

Figure 4: Preferred outflow fraction as a function of extent. One curve was determined from figure 3 and the other was determined from the simple approximation of eq. 12. The curves are nearly indistinguishable except for small mass fluxes.
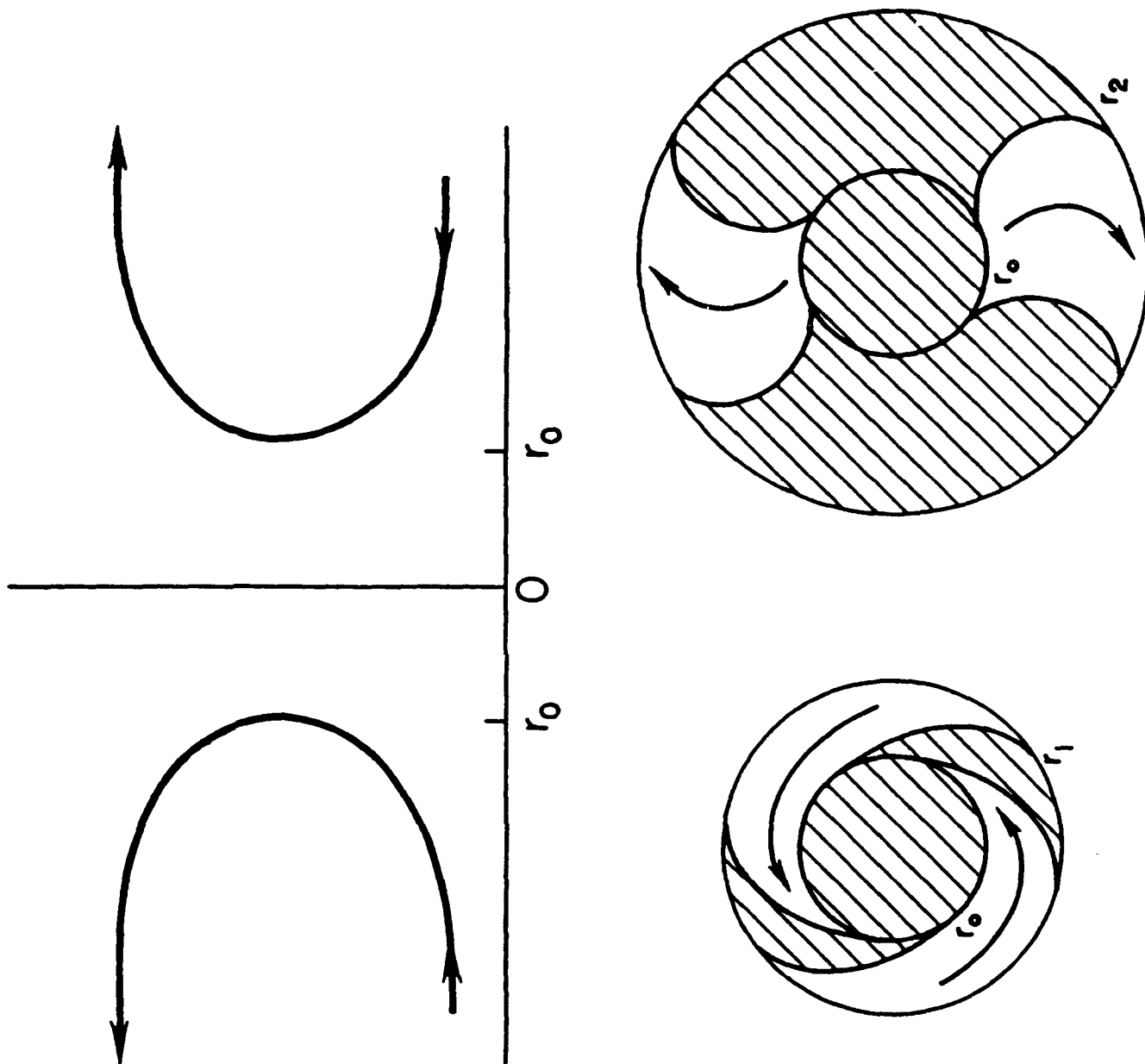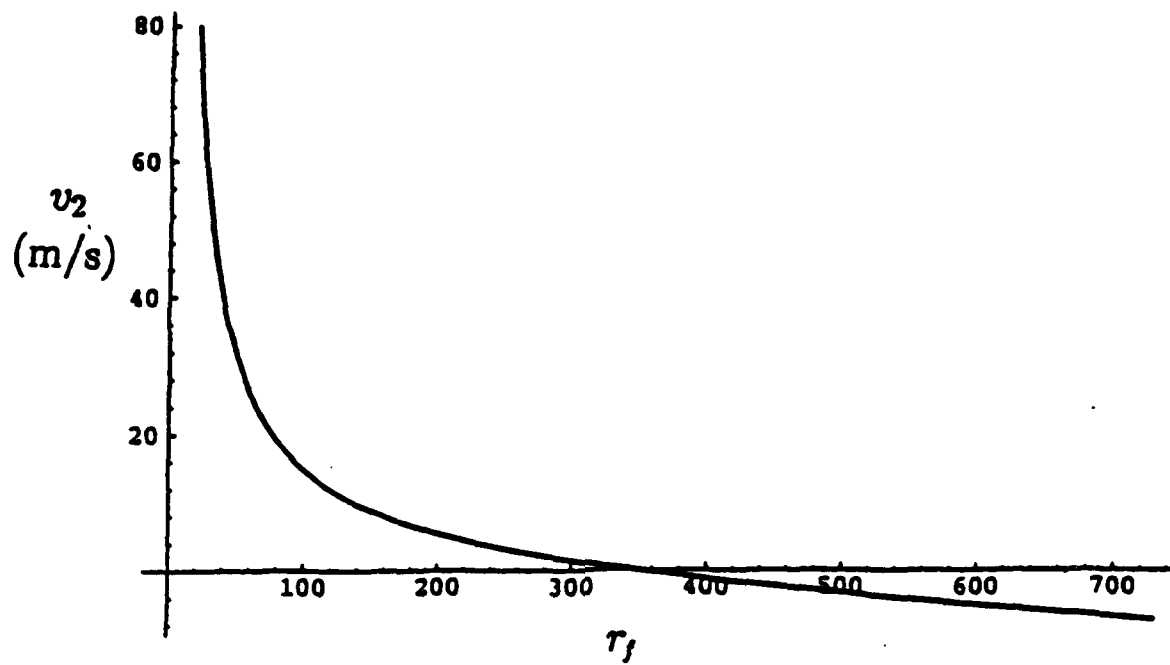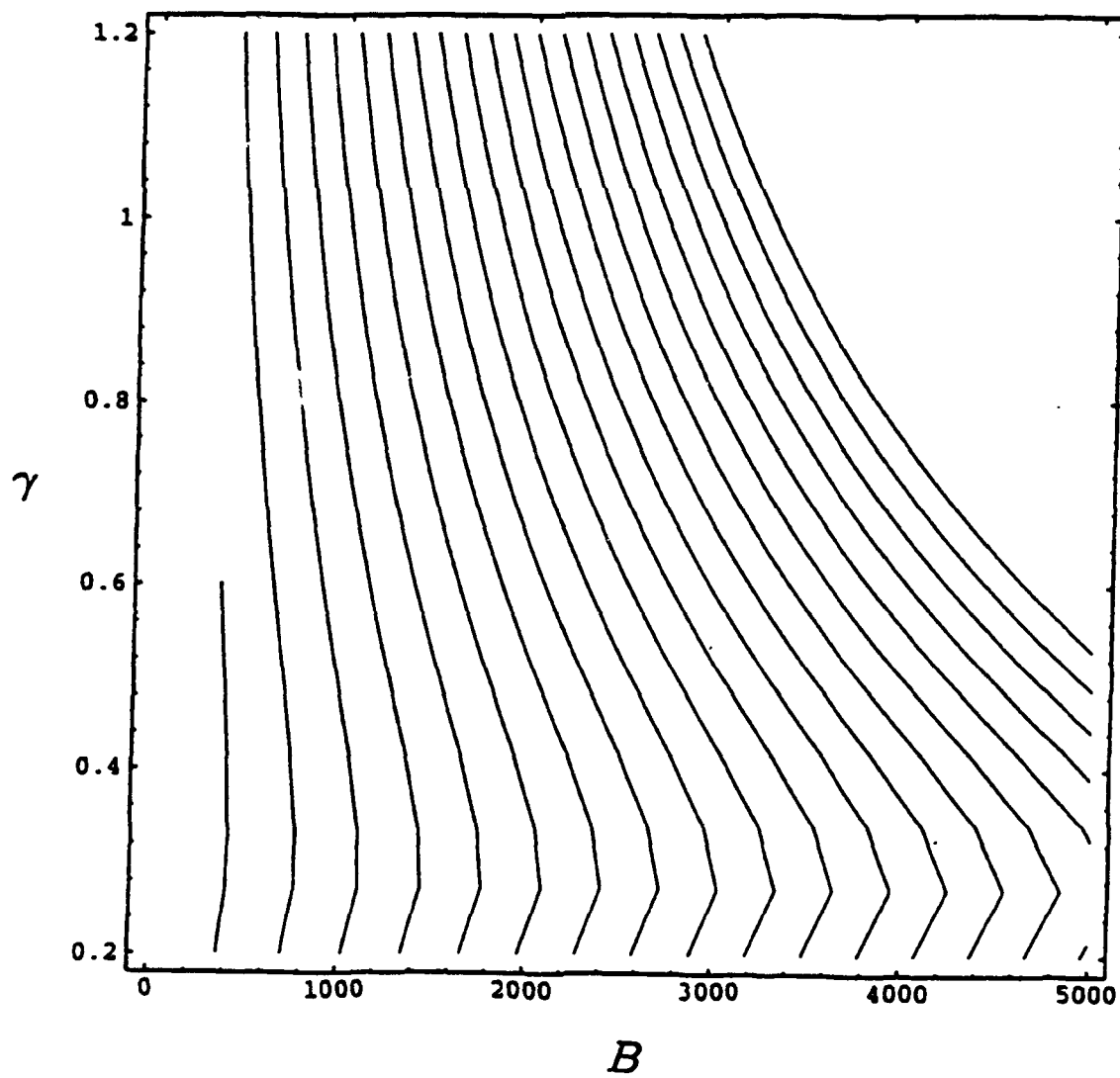
FIGURE 1

151
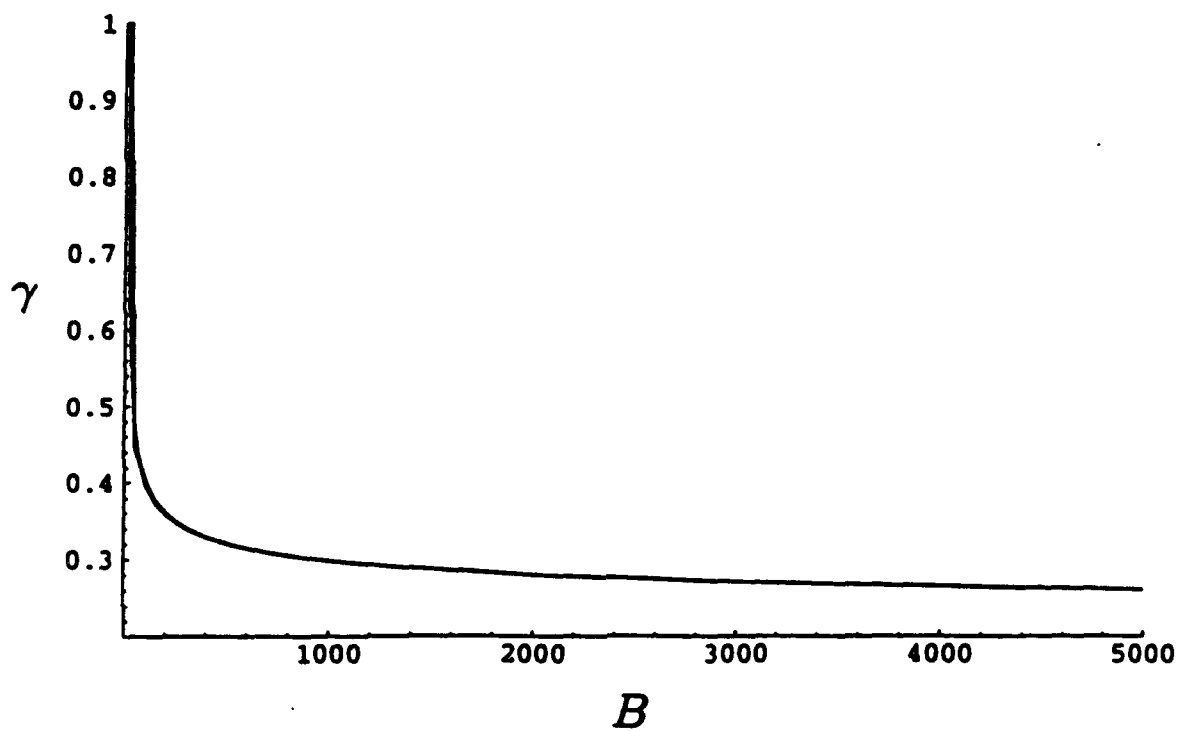
FIGURE 2

B

FIGURE 3
152

FIGURE 4

1991–1992 SCEEE Geophysics Scholar Program

Sponsored by

Department of the Air Force
Phillips Laboratory / Geophysics Directorate

Conducted by the

Southeastern Center for
Electrical Engineering Education

Final Report

Part III. Other Tropical Cyclone Papers

Prepared by: Dr. Mark David Handel
Research Location: Phillips Laboratory / Geophysics Directorate
Phillips Laboratory Research Contact: Dr. J. William Snow
Date: May 29, 1992
Contract Number: F19628-86-C-0224

TROPICAL CYCLONE INTENSIFICATION.
PART I: IMPEDIMENT TO LINEAR GROWTH

Mark David Handel

SCEEE Geophysics Scholars Program

Phillips Laboratory /GPAS

Hanscom AFB, Massachusetts 01731

U.S.A.

December 13, 1991

ABSTRACT

One of the major difficulties in tropical cyclone theory is explaining the infrequency of these intense storms. It has long been known that they depend on fluxes of heat and water vapor from the ocean surface for their primary source of energy. The same can be said of tropical cumulus clouds. A tropical cyclone, however, is only observed to occur when a fairly intense disturbance, started with a different energy source, is found over the oceans (though this is not a sufficient condition). There appears to be a threshold for whether or not disturbances intensify, both in the atmosphere and in numerical models. An analytic theory is provided to examine this threshold. In Part I, the full system of equations for our model is developed and a linearization of that system is examined to show that the growth of infinitesimal disturbances is not expected. In Part II, it will be shown how finite amplitude disturbances can intensify under conditions where the conditions for linear instability are not met.

It is assumed that the boundary layer is moist convectively unstable with respect to the air aloft and that deep convection occurs where there is Ekman pumping out of the boundary layer, the primary assumption of CISK (Conditional Instability of the Second Kind). In the model, the primary impediment to linear growth is frictional, though coupling to the descent region also increases the requirements for a growing disturbance. Quite small values of the frictional coefficients are able to stabilize a base state with large amounts of vertical available potential energy. This makes linear instability from infinitesimal perturbations unlikely, even for generous estimates of the available vertical potential energy in warm regions of the tropics in summertime.

1. Introduction

Tropical cyclones are amongst the most damaging phenomena in the atmosphere. They are capable of great intensification into the extreme examples known either as hurricanes or typhoons. Despite the scrutiny under which these disturbances have come, surprisingly little recent work has been done on the intensification process using analytic methods. This work analyzes the threshold for the development of intense tropical cyclones from, the far more frequent, weaker tropical disturbances. Here, in Part I, a linear stability analysis is performed to show that hurricane-like growth from inifinitesimal disturbances is unlikely for even generous estimates of the vertical potential energy present in the warm regions of the summertime tropics. In Part II (Handel 1991), the techniques of weakly finite amplitude stability theory are used to show how finite amplitude disturbances can intensify under conditions where the conditions for linear instability are not met. This work is largely an extension of that by Ooyama (1969).

In 1964, Ooyama, and Charney and Eliassen put forward theories for the development of tropical cyclones involving "cooperation" between the gross circulation of the cyclone and the smaller scale moist convection. Charney and Eliassen named the process Conditional Instability of the Second Kind (CISK). The energy source was the same as for moist convective instability, which implicitly was the first kind of conditional instability. The dynamics, however, were quite different. Their major contribution was to show the consequences of mass convergence in the frictional surface boundary layer organizing convection on a large scale; this is the central property of CISK. The organized convection is, in turn, responsible for the organized inflow. This inflow drives a concentration of angular momentum. Convection is far more efficient at intensifying a vortex through vorticity import than through direct heating (Schubert *et al.* 1980). These authors recognized that the perturbation must

be finite amplitude to dominate the small scale convections, which have the fastest growth rates in the simplest analyses, but they were unable to treat this aspect of the problem. It has been suggested that CISK type theories fail because as linear instability theories they imply that tropical cyclones whould be far more frequent and weaker (Emanuel 1986). We will show that this suggestion is incorrect.

Tropical cyclones are warm core disturbances over their entire depth. They form only over large bodies of water. The primary energy source is heat from the upper layer of the ocean transferred to the atmosphere in the form of latent heat, by evaporation, and later released aloft through condensation. Some of this transfer occurs before the start of the cyclone and is reflected in the potential instability of boundary layer air with respect to the air aloft. The storms are of such great intensity, and the core is of such great relative warmth, that they cannot be formed by simply lifting a set of parcels from the boundary layer under normal conditions. So, part of the development process requires increasing the entropy of the boundary layer.

This dynamical description of a tropical cyclone might then lead to the inclusion of some systems that are not tropical. In particular, polar lows may be dynamically quite similar to hurricanes (Rasmussen 1979, Emanuel and Rotunno 1989, and the works cited therein; though as discussed in Rasmussen and Lystad 1987, they disagree on the mechanism responsible for development). Midlatitude baroclinic cyclones can also extract heat from the ocean. In several well studied cases (e.g. Gyakum 1983, Bosart and Lin 1984, Reed and Albright 1986) very rapid intensification of such disturbances has been observed when they move or form over the ocean. Gyakum (1983) explicity calls on a CISK-like mechanism to explain development. The theory presented in this work for tropical cyclones may also apply to these other types of storms in various degrees.

1.1. Existence of a Threshold

We will distinguish between genesis and intensification. The creation of a large, organized, initial disturbance, by any means, will be referred to as *genesis*. Many such disturbances occur or move over the oceans. Some of these are called easterly waves, others apparently begin as instabilities of the intertropical convergence zone or along the southern monsoon shear line. A small fraction *intensify*, apparently by organizing deep convective activity, forcing unsaturated air up and out of the boundary layer so that it saturates and becomes convectively unstable, and by extracting heat from the ocean surface; all leading to the development of a very warm core. The processes leading to genesis will, on the whole, be ignored in this work, which concentrates on intensification.

At least as far back as the 1930s, the *infrequency* of tropical cyclones was vexing to those attempting to develop an explanation for these intense storms (see the discussion following Durst and Sutcliffe 1938). Hurricanes do not regularly erupt like the frequent tropical cumulus clouds, or even the occasional cumulonimbus. In the course of a year, there are thousands of cloud clusters over the tropics (McBride 1981). Few of these will grow to be classified as tropical depressions. Only a small fraction of the tropical depressions will become hurricanes (about a third, based on N. Frank and Hebert 1974). Of those that reach tropical storm status, about two-thirds will reach hurricane intensity (Gray 1975), for a global total of usually less than one hundred. Palmén (1956) noted "In spite of the fact that the necessary climatological-geographical conditions for the formation of tropical cyclones prevail over large areas of the earth during certain seasons, the actual appearance of cyclones of hurricane strength is a relatively rare phenomenon." Contrary to this, we will show that the climatological conditions for tropical cyclone development through linear instability are not met.

It is not simple to distinguish between cases where strong cyclones develop out of tiny perturbations (linear instability) and those where a sizeable perturbation is required. If one had complete and perfect data over the lifetime of a disturbance, the order of relative importance of the dominant energy source terms would change at the threshold, if one existed. Such a data set does not exist, nor anything even close to it. However, for at least some disturbances, the Atlantic easterly waves, the disturbances begin in a region of barotropic instability over land, then intensify over the water in a region with little barotropic or baroclinic energy supply (Reed and Recker 1971, Burpee 1972), indicative of a change in energy source.

1.2. Solution Procedures

The theory presented here is based on the CISK type model of Ooyama (1969). We will examine Ooyama's general system, with the inclusion of dissipation, up to second order in an amplitude expansion. The basic feature that deep convection and net heating aloft are in phase with mass convergence in the boundary layer is retained. This system can exhibit a disturbance amplitude threshold for growth. The threshold is tied to overcoming dissipation and the faster than linear growth of some of the forcing terms.

We will assume that cyclone scale oscillations are not an essential part of the intensification process. Asymmetric structures may be an important part of the environment that affects intensification, but these will not be considered necessary for intensification, though some workers have taken a contrary position (e.g. Molinari and Vollaro 1989, Challa and Pfeffer 1990, Merrill 1988). Asymmetries in the upper level will only be considered helpful in the development process to the extent that they encourage momentum mixing aloft within the anticyclone, but not as an "outside" influence.

The weakly nonlinear method has several well defined steps. The steps related to linear systems will be examined here in Part I. The mathematical procedures are described in greater detail in Handel 1990. After laying out the basic equations, the system is linearized and the resulting eigenvalue problem solved. The first order linear system is homogeneous. Only the stationary case result is needed, otherwise the solution would be dominated by the first order time dependence (though cf. Orszag and Patera 1983 where a second order expansion was performed on slowly decaying solutions). Moist convection provides a major complication in even linear problems. The thermodynamics of updrafts and downdrafts are different, so matching conditions at the dividing location(s) must be met in addition to the boundary conditions (see Haque 1952 for perhaps the first example of this). The domain location separating these regions must be solved for as an additional eigenvalue. The inclusion of dissipation, which is necessary for nontrivial stationary solutions, leads to much higher order equations than in the earlier CISK theories. Momentum diffusion also greatly increases the value of the forcing parameter needed for linear instability such that the ambient vertical potential energy even in the most favorable parts of the tropics is insufficient for growth from infinitesimal perturbations.

2. Physics of the Model

The model used is set in cylindrical coordinates (r, z), where r and z are the radius and height. The disturbance is centered and fixed on the origin of the coordinate system. All fields are assumed to be axisymmetric. The coordinate system is rotating and the Coriolis parameter, f, is constant. Essentially, the model has two layers. A third layer at the lower boundary is treated separately, but is of fixed height. The layers are designated 0 for the boundary layer, 1 for the lower layer, and 2 for the upper layer. Azimuthal (or tangential) motion; i.e. the velocities v_0, v_1, and v_2; will

be referred to as the primary circulation. Motion in any (r, z)-plane, is the secondary circulation.

We will follow Ooyama 1969 closely in setting up the basic equations, though not at all in solving them. The three layers have only two different densities, with the lower layer and the boundary layer having the same density, defined as $\rho_1 = \rho_0$, $\rho_2 = \epsilon \rho_0$, with $\epsilon < 1$. This assumption is made strictly to keep the problem as simple as possible. The *inward* radial mass flux is defined by $\psi_j \equiv -h_j u_j r \rho_j / \rho_0$, where h_j is the layer thickness and u_j is the (positive outward) radial velocity. The layout is illustrated in figure 1. In general, subscripts will always designate the layer to which the variable applies, with a generic subscript of j indicating that the expression applies to all layers.

2.1. Mass and Momentum Equations

Continuity of mass is maintained in each layer. The boundary layer is assumed to have constant thickness with respect to both space and time, so $\frac{\partial h_0}{\partial r} \equiv 0$ and $\frac{\partial h_0}{\partial t} \equiv 0$. The resulting continuity equations are:

$$0 = \frac{\partial \psi_0}{r \, \partial r} - w \tag{2.1}$$

$$\frac{\partial h_1}{\partial t} = \frac{\partial \psi_1}{r \, \partial r} - Q + w \tag{2.2}$$

$$\epsilon \frac{\partial h_2}{\partial t} = \frac{\partial \psi_2}{r \, \partial r} + Q \tag{2.3}$$

$w \equiv w^+ - w^- \equiv$ vertical velocity out of boundary layer
   $[w^+$ (upward) ≥ 0, w^- (downward) $\geq 0]$

$Q \equiv Q^+ - Q^- \equiv$ interlayer mass flux (velocity units)
   between layers 1 and 2
   $[Q^+$ (upward) ≥ 0, Q^- (downward) $\geq 0]$.

The vertical velocity out of the boundary layer, and the interlayer mass flux from latent heat release and diabatic processes, are separated into upward and downward

motions, both of which are defined to be nonnegative. There can be both upward and downward interlayer mass fluxes at the same radius, simultaneously. This indicates that there are parameterized processes on scales smaller than those resolved.

The equations for angular momentum, $M_j \equiv v_j r + f r^2 / 2$, are

$$\frac{\partial}{\partial t}(h_0 M_0) = \frac{\partial}{r \, \partial r}(\psi_0 M_0) - Z_{01} + Z_s \tag{2.4}$$

$$\frac{\partial}{\partial t}(h_1 M_1) = \frac{\partial}{r \, \partial r}(\psi_1 M_1 + \Lambda_1) - Z_{12} + Z_{01} \tag{2.5}$$

$$\epsilon \frac{\partial}{\partial t}(h_2 M_2) = \frac{\partial}{r \, \partial r}(\psi_2 M_2 + \Lambda_2) + Z_{12} \tag{2.6}$$

where the Λ_j represent the lateral eddy fluxes of angular momentum, and Z_s, Z_{ij} represent the vertical fluxes of angular momentum. In the boundary layer, lateral diffusion has been assumed to be much less than the surface stress and is ignored. The vertical fluxes are

$$Z_s = -\tau_s r / \rho_0, \qquad \tau_s = \rho_0 (k_s v_0 + C_D |v_0| v_0) \tag{2.7}$$

$$Z_{01} = M_0 w^+ - M_1 w^- + \nu (M_0 - M_1) \tag{2.8}$$

$$Z_{12} = M_1 Q^+ - M_2 Q^- + \mu (M_1 - M_2), \tag{2.9}$$

where τ_s is the surface stress, k_s is a linear drag coefficient with velocity units, and C_D is a dimensionless quadratic drag coefficient. Interlayer friction is included with the linear coefficients μ and ν. The lateral eddy flux of angular momentum is assumed to follow a simple Fickian diffusion law with respect to solid body rotation:

$$\Lambda_j = \epsilon_j \lambda_j h_j r^3 \frac{\partial}{\partial r}\left(\frac{v_j}{r}\right) = k_j r^3 \frac{\partial}{\partial r}\left(\frac{v_j}{r}\right), \tag{2.10}$$

where λ_j is an eddy diffusivity, $k_j \equiv \epsilon_j \lambda_j h_j$ is a layer integrated diffusion coefficient, and $\epsilon_1 = 1$, $\epsilon_2 = \epsilon$.

The pressure field will be assumed to be hydrostatic, so

$$p_0(r, z) = g\rho_0(h_0 + h_1 + \epsilon h_2 - z), \quad 0 \leq z < h_0$$

$$p_1(r, z) = g\rho_0(h_0 + h_1 + \epsilon h_2 - z), \quad h_0 < z \leq h_0 + h_1 \tag{2.11}$$

$$p_2(r, z) = g\epsilon\rho_0(h_0 + h_1 + h_2 - z), \quad h_0 + h_1 \leq z \leq h_0 + h_1 + h_2 ,$$

where g is the accleration due to gravity. Since homogeneous layers are assumed, the geopotential perturbations are $\phi_j(r) \equiv \left[p_j - p_j^{(0)}(z)\right]/\rho_j$, where the superscript (0) refers to values for the state of no motion. The lowest order balance in a tropical cyclone is the gradient wind:

$$\left(f + \frac{v_j}{r}\right) v_j = \frac{\partial\phi_j}{\partial r} . \tag{2.12}$$

This diagnostic constraint provides the link between development of the pressure and tangential velocity fields. The changes in these fields, however, are controlled more by the secondary circulation. From eqs. (2.11), the definition of the ϕ_j, and (2.12), it is then clear that $v_0 = v_1$ and $M_0 = M_1$. Hence, the term proportional to ν in eq. (2.8) vanishes and v_1 can replace v_0 in the stress law of eq. (2.7).

Substituting the expressions for the vertical flux of angular momentum, (2.7)–(2.9), into the equations for conservation of angular momentum, (2.4)–(2.6), yields

$$h_0 \frac{\partial v_1 r}{\partial t} = \zeta_1 \psi_0 - \frac{\tau_s r}{\rho_0} \tag{2.13}$$

$$h_1 \frac{\partial v_1 r}{\partial t} = \zeta_1 \psi_1 + (Q^- + \mu)(v_2 - v_1)r + \frac{\partial\Lambda_1}{r\,\partial r} \tag{2.14}$$

$$\epsilon h_2 \frac{\partial v_2 r}{\partial t} = \zeta_2 \psi_2 + (Q^+ + \mu)(v_1 - v_2)r + \frac{\partial\Lambda_2}{r\,\partial r} . \tag{2.15}$$

The $\zeta_j \equiv f + \frac{\partial r v_j}{r\,\partial r}$ are the total vorticity and $\zeta_0 = \zeta_1$.

Combining the equation for continuity in the boundary layer (2.1), with a radial derivative of the equation for conservation of angular momentum (2.13) yields an

expression for the vertical pumping out of the boundary layer,

$$w = \frac{\partial}{r\,\partial r}\left[\frac{k_s r v_1 + C_D r |v_1| v_1 + h_0 \frac{\partial r v_1}{\partial t}}{f + \frac{\partial r v_1}{r \partial r}}\right] . \tag{2.16}$$

The terms due to friction (those proportional to k_s or C_D) are the Ekman pumping, though we may occasionally refer to the whole of the mechanically forced vertical velocity under this rubric. For the lowest order solutions, which will be linear, the stress is then linearly proportional to the tangential velocity and the Ekman pumping is proportional to the relative vorticity. Ekman pumping and radial profiles of vertical velocity are discussed further in Part III (Handel 1992).

2.2. Diabatic Processes

Parameterization of the vertical mass flux in response to heating[1] and cooling is one of the more crucial, yet *ad hoc*, parts of this model. There exist far more sophisticated schemes for parameterizing cumulus convection than the one used here, but on the whole they are not yet amenable to analytic treatment by most mortals.

The positive vertical interlayer mass flux is taken such that $Q^+ = \eta w^+$, where η is a nonnegative entrainment parameter. This assumes that cumulus heating and entrainment only occur where there is Ekman pumping out of the boundary layer. As already noted, this is the central assumption of CISK. In a mature hurricane there are often convective bands outside of this region. However, in weak disturbances there is much less convection outside of the central region and what convection there is tends to be scattered. Scattered convection will not have any great effect on the organized lateral mass flow of the cyclone. Since we will not allow both w^+ and w^- to be nonzero simultaneously, it will be easier to include the requirement that Q^+

---

[1]Following Schubert and Hack (1982), by *heating* we refer to $\frac{d\theta}{dt}$ and by *warming* to $\frac{\partial \theta}{\partial t}$, where θ is the potential temperature.

vanish when $w^+ = 0$ in the definition of η rather than the definition of Q^+. So, we set

$$Q^+ = \eta w , \qquad (2.17)$$

with

$$\eta = \begin{cases} 1 + \dfrac{\chi_0 - \chi_2}{\chi_2 - \chi_1} , & w > 0 \quad (w^+ > 0) \\ 0 , & w \le 0 \quad (w^+ = 0) \end{cases} \qquad (2.18)$$

$\chi_0 \equiv \theta_e$ of boundary layer

$\chi_1 \equiv \theta_e$ of lower layer

$\chi_2 \equiv \theta_e^*$ of upper layer ,

where θ_e is the equivalent potential temperature and θ_e^* is the saturation equivalent potential temperature. A schematic of the parameterization is included in figure 1.

It is assumed that saturated cloud air has nearly the same potential temperature (θ not θ_e) as its environment when it reaches the upper layer, and that the χ_j mix linearly. The amount of lower layer air entrained is the amount that neutralizes the boundary layer air with respect to the upper layer. The buoyancies, or potential temperatures, of two parcels are approximately equal when their saturation equivalent potential temperatures are equal, which ignores the effects of water in all of its phases. For a saturated parcel, $\theta_e = \theta_e^*$ and there is little difference between these two measures at high altitude. Cloud air is then neutral with respect to the upper layer when its θ_e (equal to $[\chi_0 + (\eta - 1)\chi_1]/\eta$ by the mixing assumptions just given) matches θ_e^* of the upper layer (equal to χ_2). This assumes that most entrainment is lateral rather than at cloud top.[2] The only case we will be concerned with is where

---

[2]Entrainment and mixing are not equivalent, though the distinction is avoided here. There can be large amounts of entrainment and still have undilute parcels reaching cloud top. Further, it can be the case that though most of the entrainment is lateral, parcels reaching the cloud top mix mostly with other parcels at the upper reaches of the cloud.

$\eta > 1$ and $\chi_0 > \chi_2 > \chi_1$. In this case the two main layers are stably stratified and the boundary layer is conditionally unstable.

With use of this scheme there is no prognostic thermodynamic equation.[3] Even though it is assumed that there is sufficient entrainment to neutralize boundary layer parcels with respect to the upper layer, this situation is not "neutral". There is a real vertical instability at work here. It is only that the evolution of the temperature field is slow and the vertical accelerations, *per se*, are irrelevant to the problem. Rather, it is the symmetric entrainment caused by the heating and vertical acceleration that is crucial. This is consistent with the analysis of Schubert *et al.* (1980) showing that a cloud cluster is far more efficient in spinning up a vortex through its vorticity import than through direct heating.

To spin up a tropical cyclone, angular momentum must be transported inward throughout a fairly thick layer (about 5 km). The magnitude of $\chi_0 - \chi_2$, found in the numerator of eq. (2.18), is a measure of the convective potential energy available to drive the system. To advect, entrain, and then vertically stretch a parcel from the lower layer, the static stability between the layers must be overcome. Hence, decreasing $\chi_2 - \chi_1$, found in the denominator of eq. (2.18), increases the forcing parameter. Increasing χ_0 in the boundary layer only increases the numerator in η. Decreasing χ_2 by cooling the upper layer does double duty.

Once released, an unstable parcel will rise, release heat through condensation, and accelerate upwards. The acceleration and heat release will induce an accompanying solenoidal circulation. Fluid that flows inward toward the path of the rising parcel and follows it upward should be considered as entrained, though no mixing has occurred.

[3]A closely related scheme was recently used by Fraedrich and McBride (1989), while examining the CISK model of Charney and Eliassen, but who seem to be under the misimpression that elimation of a prognostic thermodynamic equation is similar to a lack of reliance on a large reservoir of convective potential energy. The dynamics of their "free-ride" balance are essentially the same as that of Ooyama's parameterization from twenty-five years earlier, used here.

In a hydrostatic model it is impossible to close the secondary circulation of a steady solution without some diabatic cooling process. The simplest parameterization of such a process is Newtonian radiative cooling. Cooling is set proportional to the difference between the mean potential temperature with no motion and the mean potential temperature of the two layers, θ_m:

$$Q^- = \beta(\theta_m - \theta_m^{(0)}) , \tag{2.19}$$

where β is a cooling constant in units of $m\,sec^{-1}K^{-1}$ and $\theta_m^{(0)}$ is θ_m for the rest state. Between two given pressure surfaces, the mean temperature is proportional to the thickness between the surfaces:

$$\theta_m - \theta_m^{(0)} = \frac{\phi_2 - \phi_1}{c_p(\pi_1 - \pi_2)} , \tag{2.20}$$

where $\pi_j \equiv (p_{0j}/p_{00})^{2/7}$, c_p is the specific heat of dry air at constant pressure, and p_{0j} is an average pressure for the layer j. Making use of the gradient wind relation, eq. (2.12),

$$\frac{\partial \theta_m}{\partial r} = \frac{\left(f + \frac{v_2}{r}\right)v_2 - \left(f + \frac{v_1}{r}\right)v_1}{c_p(\pi_1 - \pi_2)} . \tag{2.21}$$

This assumes that $\frac{\partial \pi}{\pi \partial r} \ll \frac{\partial \phi}{\phi \partial r}$ or that $\frac{2}{7}\frac{p - p^{(0)}}{p} \ll 1$, which holds. This also assumes that $\theta_m = \theta_m^{(0)}$ where $\phi_2 = \phi_1$. We then obtain an expression for the radial derivative of the cooling, though not the cooling directly:

$$\frac{\partial Q^-}{\partial r} = b\left[\left(f + \frac{v_2}{r}\right)v_2 - \left(f + \frac{v_1}{r}\right)v_1\right] \tag{2.22}$$

where $b \equiv \beta/[c_p(\pi_1 - \pi_2)]$. Radiative cooling is allowed everywhere in the domain. A shortcoming of this scheme is that there is no difference in cooling rates between cloudy and cloud free regions (see W. Frank 1977 for observational evidence of the importance of this difference). In real tropical cyclones the decreased radiative cool-

168

ing in cloudy regions may encourage intensification, though in the scheme used here the cooling is strictly dissipative.

We will also need to determine the variations of the χ_j that affect both Q^+ and Q^-. In the boundary layer

$$\frac{\partial \chi_0}{\partial t} + u_0 \frac{\partial \chi_0}{\partial r} + \frac{w^-}{h_0}(\chi_0 - \chi_1) = \tag{2.23}$$
$$\frac{C_\varepsilon}{h_0}|v_1|(\chi_s - \chi_0) + \frac{\partial}{r \partial r}\left(k_\chi r \frac{\partial \chi_0}{\partial r}\right) ,$$

where C_ε is a bulk exchange coefficient and k_χ is a diffusion coefficient. The equivalent potential temperature of a parcel saturated at the sea surface temperature is designated χ_s. This value was kept fixed in the weakly nonlinear calculations, though for strong disturbances it must be taken as a function of surface pressure and would provide an additional boost for intensification.

The value of χ_1 will be assumed constant and there is no turbulent entropy exhange between the boundary and lower layers in eq. (2.23). This was done mostly for simplicity and in retrospect may have been less than wise. As will be seen in Part II, a consequence of this is unrealistic growth of χ_0 for realistic values of the bulk transfer coefficient C_ε.

There are also changes aloft. The first order pressure changes are the result of changes in the thickness of the layers, not changes of the potential temperature or density within each layer. At next order the atmosphere must be warming up due to the latent heat release, which is tied to the change of layer thickness. In the upper layer the perturbations will be approximated as

$$\chi_2 = \chi_2^{(0)} + \Xi(\theta_m - \theta_m^{(0)}) . \tag{2.24}$$

This assumes that changes in the saturation equivalent potential temperature are Ξ times the perturbations of the potential temperature in the upper troposphere. We will set $\Xi = 1.2$. This ratio is also approximately equal to ratio of the dry adiabatic

lapse rate to the moist adiabatic lapse rate in the upper troposphere. The mean temperature perturbation is as determined by eq. (2.20). The assumption that most of the mean column temperature perturbation is in the upper layer is supported by observations such as those shown in Hawkins and Rubsam 1968.

3. Linear Theory

The inviscid linear analysis presented in Ooyama 1969 is an inappropriate starting point for weakly nonlinear analysis since it admits no stationary solutions with motion in the lower layer. It also fails to consider the differences in the moist physics between regions with ascent and those with descent. The stationary system with friction will be the basis of the finite amplitude expansions in Part II, though we will briefly examine the linear time dependent case here. Examination of this linear system will show why the growth of hurricane-like disturbances from inifinitesimal disturbances is unlikely. A major complication in performing a linear analysis of this system is the difference in the dynamics of updrafts and downdrafts, resulting in a double eigenvalue problem.

3.1. Expansions

The fields will be expressed with an amplitude expansion. The azimuthal velocity in each of the layers is represented as

$$v_j = a v_j^{(1)} + a^2 v_j^{(2)} + \cdots \tag{3.1}$$

for $j = 0$, 1, 2. The nondimensional amplitude, a, is a Rossby number measured as $a \equiv V/(fL)$, where V is the magnitude of the velocity perturbation and L is the length scale (as yet to be determined). We will only be examining solutions that are cyclonic in the lower layer near the axis, so both a and V are nonnegative. The

nondimensional amplitude will be assumed to be much less than one, while the radial structure functions $v_j^{(1)}(r)$ are for the time being dimensional and will be normalized to have maxima of order fL. The amplitude of stationary solutions will have the symbol $\bar{a}$ and the amplitude of infinitesimal time dependent perturbations will be designated $\tilde{a}$. All of the fields of motion will be expanded in similar fashion. Superscripts in parentheses will indicate the expansion order of a variable or parameter. Some parameters have constant values in the absence of motion; these will be designated with a superscript (0).

It will also be necessary to expand the forcing parameter, η, and the growth rate, γ, which are both scalers. The parameter η serves as the forcing parameter of the system. At lowest order

$$\eta^{(0)} = 1 + \frac{\chi_0^{(0)} - \chi_2^{(0)}}{\chi_2^{(0)} - \chi_1^{(0)}} \; . \tag{3.2}$$

For stationary solutions, $\gamma^{(0)} = 0$ and $\gamma^{(1)} = 0$; for slowly varying solutions $\gamma^{(0)} = 0$ and $\gamma^{(1)} \neq 0$. To determine the stability boundaries and perform the weakly nonlinear analyis, only stationary and slowly varying equations are needed. Some limited results from the linear time dependent system will be presented below, but the equations will not be examined here (they can be found in Handel 1990).

At first order for stationary or slowly varying solutions, the vertical velocity expands to

$$w^{(1)} = \frac{k_s}{f} \frac{\partial r v_1^{(1)}}{r \, \partial r} \; ; \tag{3.3}$$

the Ekman pumping is proportional to the low level vorticity. The cumulus heating expands as

$$Q^{+(1)} = \eta^{(0)} w^{(1)} = \eta^{(0)} \frac{k_s}{f} \frac{\partial r v_1^{(1)}}{r \, \partial r} \tag{3.4}$$

and the derivative of the radiative cooling as

$$\frac{\partial Q^{-(1)}}{\partial r} = bf(v_2^{(1)} - v_1^{(1)}) .$$

(3.5)

3.2. Linear System

One could directly reduce the resulting linear problem to an eighth order equation in one variable, though it then becomes very difficult to follow the physics. This also requires taking several derivatives of the discontinuous parameter η, which is very problematic when it comes time to impose matching conditions at the boundaries between ascent and descent regions. There is a clearer way to proceed using a vector equation. Two new variables, proportional to the radial flow, are now introduced: $v_3 \equiv (f/k_s)\psi_1/r$, $v_4 \equiv (f/k_s)\psi_2/r$. The constants k_s and f are included in the definitions to make v_3 and v_4 have the same units (velocity) as v_1 and v_2. Neither of those constants will ever be set to zero, since that would eliminate the basic driving physics of the problem.

The linear stationary versions of the momentum equations (2.14) and (2.15) at lowest order are

$$0 = k_1^{(0)} r \square^2 \bar{v}_1^{(1)} + \mu(\bar{v}_2^{(1)} - \bar{v}_1^{(1)})r + k_s \bar{v}_3^{(1)} r$$

(3.6)

$$0 = k_2^{(0)} r \square^2 \bar{v}_2^{(1)} + \mu(\bar{v}_1^{(1)} - \bar{v}_2^{(1)})r + k_s \bar{v}_4^{(1)} r .$$

(3.7)

The first term in each of these represents lateral diffusion of angular momentum, the second term is the vertical diffusion, and the last term is the lateral advection of planetary vorticity by the secondary circulation (though the factor f is hidden in the definitions). The overbars indicate stationary solutions. The box operator is defined as $\square^2 v \equiv \frac{\partial}{\partial r}\left(\frac{\partial r v}{r \partial r}\right)$. Though the operator has been defined in terms of partial derivatives, the same notation will be used when full derivatives with respect to r are intended.

172

The radial derivative of the first order stationary version of the continuity equations (2.2); with the first order expansion of w, eq. (3.3); the substitution of the heating parameterization, eq. (3.4); the cooling parameterization, eq. (3.5); and multiplication throughout by $r k_2^{(0)} f / k_s$; is

$$0 = k_2^{(0)} r \Box^2 \bar{v}_3^{(1)} - k_2^{(0)} r \frac{d}{dr} \left[(\eta^{(0)} - 1) \frac{d r \bar{v}_1^{(1)}}{r \, dr} \right] + \frac{k_2^{(0)} b f^2}{k_s} (\bar{v}_2^{(1)} - \bar{v}_1^{(1)}) r \; . \qquad (3.8)$$

The first term here is the derivative of the lateral flow divergence, the next terms are the derivatives of the vertical transport to the upper layer and the vertical transport from the lower layer, and the last term is the derivative of the descent into the lower layer from radiative cooling.

Rather than use the upper level continuity equation (2.3), we use the sum of the continuity equations. Integrating once, with the boundary condition of no flow at the origin, yields

$$\bar{v}_1^{(1)} + \bar{v}_3^{(1)} + \bar{v}_4^{(1)} = 0 \; . \qquad (3.9)$$

Since we will be dealing with two separate domains, the application of the boundary condition at the origin does not constrain the integral of the summed continuity equations in the outer region. The boundary condition that there is no net lateral flow at the outer boundary of the domain leads to the same result for the outer region. The three terms are proportional to the inward horizontal transport in the boundary layer, lower layer, and upper layer, respectively. The last four equations can be reduced to three, using eq. (3.9) to eliminate $\bar{v}_4^{(1)}$. This reduced system can easily be expressed in matrix form but is not written out here.

Since we will later be determining solvability conditions on inhomogeneous versions of the linear system, it is necessary to construct the adjoint system. It would be preferable, however, to transform the linear system into one that is self-adjoint.

a process that does not affect the linear solutions. The transformed, self-adjoint system for $\mathbf{v} = (v_1, v_2, v_3)$ is:

$$r\frac{d}{dr}\left(k_2^{(0)}B\frac{dr\bar{v}^{(1)}}{r\,dr}\right) + k_s Ar\bar{v}^{(1)} = 0 \qquad (3.10)$$

where

$$B = \begin{pmatrix} KM - M - BK^2 + 1 - \eta^{(0)} & M - BK & 1 \\ M - BK & -B & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$A = \begin{pmatrix} \frac{M^2}{K} - M + BK - B & -\frac{M^2}{K} + B & -\frac{M}{K} \\ -\frac{M^2}{K} + B & \frac{M^2}{K} & \frac{M}{K} \\ -\frac{M}{K} & \frac{M}{K} & \frac{1}{K} \end{pmatrix},$$

with the nondimensional parameters defined as $M \equiv \mu/k_s$, $K \equiv k_1^{(0)}/k_2^{(0)}$, and $B \equiv bf^2 k_2^{(0)}/k_s^2$.

At the center (origin), the tangential velocities and the radial transports must vanish. The same conditions will be used at the outer boundary of the domain, $r = r_1$, so

$$\mathbf{v}|_{r=0} = 0 \quad \text{and} \quad \mathbf{v}|_{r=r_1} = 0 \qquad (3.11)$$

provide all of the boundary conditions in vector form.

One of the major complications in this problem is the discontinuity in the co-efficients of the linear equations at the boundaries between updraft and downdraft regions. We will assume that the region closest to the origin has positive w and that there is only a single region outside of that, with Ekman suction throughout. By solving this system, we will show that this assumption is consistent, though it is

174

probably not unique. The internal boundary where $w = 0$ is labeled r_0, and there, at first order,

$$\left. \frac{dr v_1^{(1)}}{r \, dr} \right|_{r=r_0} = 0 \, . \tag{3.12}$$

This is required through eqs. (2.18) and (3.3). It will be necessary to find r_0 as an eigenvalue of the problem.

There must be as many matching conditions at the boundary between these two regions as the order of the system, which is the same as the number of boundary conditions imposed at the two endpoints. The required matching conditions are that all of the v_j and their first derivatives (or equivalently $\frac{dr v_j}{r \, dr}$) are continuous. These conditions can be derived rigorously by integrating eq. (3.10) across r_0 assuming only that v is finite in a small neighborhood of r_0. For matching conditions we require that the vectors

$$\mathbf{v}^{(1)} \quad \text{and} \quad \frac{dr \mathbf{v}^{(1)}}{r \, dr} \quad \text{are continuous at } r_0 \, . \tag{3.13}$$

3.3. Basis Functions and Nondimensionalization

Since $\frac{d}{dr} \left(\eta^{(0)} \frac{dr v_1^{(1)}}{r \, dr} \right) = \eta^{(0)} \Box^2 v_1^{(1)}$, because of eq. (3.12), we can replace eq. (3.10) with

$$k_2^{(0)} B \Box^2 \bar{\mathbf{v}}^{(1)} + k_s A \bar{\mathbf{v}}^{(1)} = 0 \, . \tag{3.14}$$

Both B and A are piecewise constant matrices. Therefore we can deal separately with each region as a standard linear problem. In the ascent region we will assume that the solutions are sums of first order Bessel functions so that $\Box^2 v = -m^2 v$. The m^2 then satisfy $\det \left(k_s A - m^2 k_2^{(0)} B \right) = 0$. For $m^2 < 0$, the solutions are modified Bessel functions. The simplest nondimensionalization uses a length scale of $L = \sqrt{k_2^{(0)}/k_s}$. For any limits that still allow stationary hurricane-like solutions, both

of the parameters used are kept finite. The square of the nondimensional wavenumbers (in roman type) are transformed as $\mathrm{m}^2 = (k_s/k_2^{(0)})m^2$. The nondimensional wavenumbers satisfy

$$K\mathrm{m}^6 - \left[\eta^{(0)} - 1 - M(1+K)\right]\mathrm{m}^4 + [M + B(1+K)]\,\mathrm{m}^2 + B = 0 . \qquad (3.15)$$

The degree of this expression implies that there are six linearly independent solutions to eq. (3.14) in the ascent region. The frictional dependence of the form of the linear solutions has now been greatly simplified. Rather than needing the not-well-known magnitudes of the frictional parameters, we need only their ratios. It is straightforward to derive the dispersion relation for the time dependent problem. The result is eighth order in m, second order in γ, and requires four additional nondimensional parameters: three pertaining to the layer depths and one to the reduced gravity. Some results from this will be discussed below and shown in figure 3.

If all of the roots m^2 were negative, i.e. there were no Bessel function solutions but only modified Bessel function solutions, or any of the roots are complex, it would be impossible to satisfy the homogeneous boundary conditions at the origin and the matching conditions, except with the trivial solution. These conditions provide requirements for a minimum value of $\eta^{(0)}$ which will be designated $\eta_f^{(0)}$. This in fact is the same minimum value required for the existence of linearly unstable modes in the ascent region. However, this is not necessarily the critical value for linearly unstable solutions meeting the boundary conditions and the matching conditions [see discussion on p. 27].

Some limiting cases are helpful. In the limit of small radiative cooling ($B \ll M, 1$)

$$\eta_f^{(0)} \approx 1 + M(1+K) + 2\sqrt{MK} + B(1+K)\sqrt{K/M} + BK^2/M . \qquad (3.16)$$

It is worth noting that *for believable values of the frictional coefficient ratios, this lower bound on $\eta^{(0)}$ is significantly greater than 1* ($\eta_f^{(0)} = 1$ for the inviscid case). Plots of $\eta_f^{(0)}$ are provided in figures 2. Note that in the singular limit of K → 0, then $\eta_f^{(0)} = 1 + M$ is the requirement for at least one positive value for m^2. The lower bound increases rapidly with M, though more slowly with K. This difference becomes less pronounced as one increases B. The graph does not extend to larger values of the parameters because these lead to values of $\eta_f^{(0)}$ so large as to be irrelevant.

Figure 3 shows growth rate as a function of squared wavenumber using several values of the forcing parameter, $\eta^{(0)}$, for basis functions with unconditional heating (cooling in descent regions). These functions are not solutions to the boundary value problem, which are discussed below. The growth rate is scaled as $\Upsilon = \gamma/f$. Parameter values are set at M = 0.5, K = 0.2, B = 0.005, ϵ = 0.4. The layer depths were set at $h_0 = 1\,\text{km}$, $h_1 = h_2 = 5\,\text{km}$. The curves are similar to those found in figure 3 of Ooyama 1969. However, the critical value for $\eta_f^{(0)}$ of 2.74 is much greater than that found by Ooyama of about 1.1. This is mostly due to the smaller value of k_s and the larger value of μ chosen here. Though the maximum growth rate increases faster than linearly with increase in $\eta^{(0)}$, when $\eta^{(0)}$ is nearly 4 the growth rate is still only about $0.05f$. This is still much slower growth than is observed for larger than observed amounts of the forcing parameter. When we consider only solutions to the full boundary value problem, the growth rates will drop even lower.

3.4. Linear Solutions

An example of a solution to the linear stationary problem is shown in figure 4. Solutions were found as two linear combinations of Bessel and modified Bessel functions for the ascent and descent regions separately. Each of the chosen Bessel functions satisified the basic linear equations in their respective regions. A double iterative scheme was used to find the basis function coefficients and the two eigenvalues for r_0

and $\eta^{(0)}$, such that the boundary and matching conditions were satisified. The eigenvalue of the forcing parameter is designated $\eta_s^{(0)}$. It is the only value of the forcing parameter compatible with boundary conditions, matching conditions, and physical parameters for a stationary solution. Due to the insensitivity of $\eta_s^{(0)}$ with respect to changes in r_1, it was not possible to fix $\eta_s^{(0)}$ and search for r_1 as an eigenvalue. Further details on the iterative scheme can be found in Handel 1990.

In the time dependent tropical cyclone related problems of say Haque (1952) and Charney and Eliassen (1964), there was no additional constraint corresponding to eq. (3.12) and a problem with only a single eigenvalue was needed. This allowed a continuous relation between the growth rate and the horizontal length scale of the updraft, with the downdraft extending out to infinity. The works of Bretherton (1987) and Emanuel *et al.* (1987) are more closely related mathematically to the problem presented here, i.e. they have different physics in two different regions *and* an additional condition leading to a double eigenvalue problem, though they are not on tropical cyclones. In those works the two parameters that were varied as eigenvalues where the divider between the regions of different physics and the growth rate. Instead of a continuous relationship, the solution set is a countable (though infinite) set of points in the parameter space of length scale and growth rate, with interest primarily in the solution with only a single updraft in the inner region.

What separates the problems with two eigenconditions from those with only one is the requirement that w vanish at r_0, rather than just change sign with a jump discontinuity. This requirement is imposed here because the problem is viscous. In the inviscid problem there can be a jump discontinuity in the vertical velocity. If a jump is allowed, there is no condition that "defines" r_0; it can vary over a range. Here, when stationary solutions are being considered, the growth rate is zero and additional parameter is still needed to be determined as an eigenvalue of the system. The problem was solved by varying $\eta^{(0)}$ with a fixed outer boundary.

For the solutions to the linear time dependent problem, used to generate figure 5, the growth rate was set and the value of $\eta_s^{(0)}$ was sought. This kept the solution procedures for the stationary and time dependent problems similar. The time dependent problem, however, was higher order and required two additional basis functions in each region.

Several basic features should be noted in the linear stationary solution. The lower layer has cyclonic circulation throughout the domain, $v_1^{(1)} \geq 0$. In the upper layer the flow is everywhere anticyclonic, $v_2^{(1)} \leq 0$. It is appropriate for a solution applying to a young disturbance not to be cyclonic at small radius in the upper layer since the air flowing outward did not begin with large amounts of angular momentum. This is also consistent with the observations of Black and Anthes (1974) that weak cyclones and ones still in early development do not have a cyclonic regime in the outflow. There is inflow throughout the entire lower layer, $v_3^{(1)} \geq 0$. The maximum strength of the anticyclone aloft is at a greater radius (about double) than that of the maximum wind of the the low level cyclone.

The vertical velocity is greatest in the center. (Linear and nonlinear drag laws lead to very different spatial distributions for the Ekman pumping, and it is a linear drag law which controls the lowest order solutions here.) The radius of maximum cyclonic winds is smaller than the radius where w crosses zero. It is also of order one nondimensionally and in Part II will be taken to be of order $L \sim 100\,\text{km}$ dimensionally.

Note that only the values of B, K, M, and r_1 are needed for the linear stationary calculation. The solution in the figure has been normalized so that the maximum of $v_1^{(1)}$ is one and the length scale is nondimensionalized with $r \equiv r/L$. Velocities in the lower and upper levels are directly comparable. The vertical velocity can be obtained from the curve for $\frac{d r v_1^{(1)}}{r\,dr}$ after scaling by $k_s/(fL) = k_s^{3/2} f^{-1} k_2^{(0)-1/2} \sim 3 \cdot 10^{-4}$ for comparison (see Part II for a discussion of parameter values). The lower level

horizontal velocity, indirectly given by v_3, must be scaled by $-(h_1 f/k_s)^{-1} \sim -10^{-2}$ for comparison.

All of the parameters will be kept fairly small. Large values of M, K, and B, would lead to unrealistically large requirements for $\eta_s^{(0)}$. We will later make use of solutions with $\eta_s^{(0)}$ of order 2. Even an $\eta_s^{(0)}$ of order 2 seems to imply a much greater amount of convective available potential energy (CAPE) than even those who believe that CAPE exists would claim. However, when the linear solutions are used in the nonlinear calculations of Part II, it will be seen that solutions with an $\eta_s^{(0)}$ of say 2.5 are relevant to finite amplitude disturbances in an environment with an ambient forcing parameter, η_a, of say 1.2, which does not require a large amount of positive buoyancy for lifted parcels.

The assumption that the vertical diffusion ratio, M, is small is equivalent to assuming that the strong surface friction exceeds internal vertical momentum mixing. In other theoretical models this is usually taken to be zero. In a mature hurricane, with large vertical convective transports, the diffusive momentum transports are of the same order. For the linear case, where the explict inclusion of cumulus momentum transports does not yet appear, the parameterized turbulent transport serves as a proxy. The values chosen for M will fall mostly in the range from 0.1 to 0.5.

For K to be small, momentum diffusion must be greater in the upper layer than in the lower layer. We will assume that this is the case. This is justified by the presence of the unresolved asymmetries and jet-like structures of the upper layer, along with inertial instabilities on short time scales. The large deviations from cylindrical symmetry create much greater internal velocity gradients than appear in the azimuthally averaged quantities. Our axisymmetric model can compensate for the lack of asymmetries aloft through a greater eddy viscosity. Both observational and modelling works (e.g. Challa and Pfeffer 1990 and the references cited therein)

have also shown that momentum mixing is stronger and more important aloft. So, the values chosen for K will also fall mostly in the range from 0.1 to 0.5.

Radiative cooling is a weak process compared to the others in this system, though still necessary to close the circulation of steady solutions. In the downdraft region, if B is made too large, oscillatory solutions would develop (some m^2 must be complex), greatly complicating the problem. We arbitrarily restrict B to prevent this, but choose it to be nearly as large as this limit.

Solutions were determined for a range of the independent parameters. Properties of these solutions are shown in table 1. It can be seen from items 7–9 and 14–18 that there is little variation in the form of the solutions as r_1 varies, nor is there much variation in the eigenvalue for $\eta^{(0)}$, except as r_1 approaches r_0. As r_1 increases, evaluation of the growing Bessel and modified Bessel functions becomes less accurate. Hence, the very weak nonmonotic variations of $\eta_*^{(0)}$ and r_0 with increasing r_1 may be an artefact of the limitations of the function evaluation algorithms and should be viewed with caution. The lack of sensitivity of $\eta_*^{(0)}$ to variation of r_1 is why it is not possible to vary r_1 in search of eigenvalues. Given a parameter set $\{M, K, B\}$, there is only a very small range of $\eta^{(0)}$ for which solutions exist. At its minimum, $\eta_*^{(0)}$ is still significantly greater than $\eta_f^{(0)}$. *This indicates that the actual critical value of $\eta^{(0)}$ for solutions consistent with the boundary conditions is greater than that which simply ensures positive real values of* m^2. For linear growth still greater values of $\eta^{(0)}$ are required., as can be seen in figure 5. The growth rate increases approximately as the 3/2 power of the supercriticality in the forcing parameter, based on the figure. For comparison with figure 3, with a forcing parameter of 3.75, the growth rate is only about $0.03f$. This is clearly insufficient to explain tropical cyclone intensification.

Decreases in M lead to increases in the anticyclone relative to the cyclone. This is consistent with the simple view that the greater the amount of vertical diffusion the lesser the amount of the vertical shear. Comparisons can be made from table 1

using item sets $\{1, 4, 20\}$; $\{2, 5, 8\}$; and $\{6, 15, 21\}$, which are presented as curves in figures 6 and 7. Increases in M lead to increases in the required forcing parameter $\eta_s^{(0)}$, decreases in the relative anticyclone strength ($R_2 \equiv |\max(v_2)/\max(v_1)|$), increases in the lower layer inflow ($R_3 \equiv |\max(v_3)/\max(v_1)|$), and decrease in the relative size of the updraft region r_0.

Variations in K may be examined with items sets from table 1 of $\{1, 2, 3\}$; $\{4, 5, 6\}$; $\{8, 15, 19\}$; and $\{20, 21\}$. These comparisons can also be seen comparing across curves at constant M in figures 6 and 7. Increases in K, which correspond to increases in the eddy viscosity of the lower layer, lead to increases in all of the required forcing, the upper level anticyclone (or weakens the lower cyclone), lower level inflow (to compensate the diffusion), and relative size of the updraft region. Since K is also proportional to the ratio of the layer depths, increasing K is a proxy for increasing the depth of the cyclone layer or decreasing the depth of the anticyclone layer. Hence it is detrimental to cyclone development to have a thin outflow layer. In actual cyclones, strong asymmetries aloft lead to increased momentum mixing in the upper layer. This increases $k_2^{(0)}$ with respect to $k_1^{(0)}$ and reduces K. Without greater mixing aloft, K would be greater than one and the required forcing parameter for linear instability would be very large.

If vertical diffusion is fairly strong, i.e. M not small, and lateral diffusion in the lower layer fairly weak, i.e. K small, the cyclone and anticyclone of the weak stationary solutions are of approximately equal strength. It is not clear in the atmosphere how much more turbulent the outflow layer is than the inflow. However, if the eddy viscosities of the two layers are similar, then $K \approx 1/\epsilon > 1$, which results in solutions with very strong anticyclones aloft, unlike any observed disturbances.

The effect of variation in radiative cooling can be seen by comparing the set of items $\{7, 10, 11\}$ in table 1 (the difference in the outer radii is not particularly

significant). The primary effects of increases in the radiative cooling are to require increased forcing and to decrease the strength of the anticyclone.

4. Discussion and Conclusions

Though some have argued that there is very little CAPE in the tropics (e.g. Betts 1982, Xu and Emanuel 1989), we will first examine the most generous estimates. These assume pseudoadiabatic ascent for parcels originating at low levels in the boundary layer. For soundings obtained during hurricane season in hurricane prone regions, the buoyancies of boundary layer parcels lifted to the upper troposphere are maximally about 8 K (Palmén 1948). Xu (1987) obtained a value of 5 K for an average summertime sounding over Truk, an area of moderate tropical cyclone activity, so slightly larger values in areas and periods of greater activity are to be expected. This provides a value for $(\chi_0^{(0)} - \chi_2^{(0)})$. Using the values for equivalent potential temperature and saturated equivalent potential temperature from Xu 1987, the stability of parcels averaged over the lower troposphere with respect to the upper troposphere is about 4 K, providing a value for $(\chi_2^{(0)} - \chi_1^{(0)})$. This yields an ambient forcing parameter value of $\eta_a = 3.0$. Given our present knowledge of the vertical structure of the atmosphere in the tropics, it would be difficult to justify a larger value for the ambient forcing parameter. With the inclusion of water loading effects and mixing processes, the appropriate value may be significantly smaller.

In the inviscid calculation of Ooyama (1969), without a coupled dry descent region, the critical value of the forcing parameter was only 1. This implied the possibility of instability if there was *any* CAPE for boundary layer parcels. In the calculation presented here, it has been shown that the critical forcing parameter is significantly greater than one. Even in the example solution fully shown, the required forcing parameter was 2.75, with the assumption that the integrated momentum

mixing throughout the troposphere was only half the boundary layer friction and the eddy diffusivity (λ not k) of the upper layer was only about a tenth of the lower layer value. For weak disturbances, which have not yet developed strong asymmetries aloft, the lateral eddy diffusivites are likely to be closer, K larger, and the required forcing parameter still greater. If the eddy diffusivities and depths were equal for the two layers, then $K = 1/\epsilon \sim 3$. As can be seen from table 1, for $K = 1.0$, with $M = 0.5$, then $\eta_a^{(0)} > 4$. From figure 5, where the stationary solution requires $\eta_a^{(0)} = 2.75$, an ambient forcing parameter of 3.0, yields an e-folding time of more than 20 days. An ambient forcing parameter of about 4.7 would be required to reduce this to the needed time of about 2 days. This leads us to the conclusion that the growth of infinitesimal hurricane-like disturbances through a linear CISK mechanism is unlikely to produce intense tropical cyclones on the time scale observed.

This is an encouraging result for use of a CISK theory of tropical cyclone development. One of Emanuel's (1986) objections to CISK is that it is a linear theory and if such linear instability actually existed "weak tropical cyclones should be ubiquitous and not confined to maritime environments", rather than strong and infrequent. Linear theory can not capture the knowledge that tropical cyclones arise out of rather strong preëxisting disturbances, such as easterly waves. However, CISK need not be restricted to a linear analysis. As will be seen in Part II, extension of the model presented here to second order allows rapid intensification under conditions shown here to be linearly stable.

Here, in Part I, linear stationary solutions were found for a system with CISK type forcing and dissipative processes, including internal friction and radiative cooling. Inclusion of friction is a singular perturbation to the equations of motion that greatly increases the thermodynamic forcing required for *linear* instability. However, our result is not as sensitive to the values of the frictional parameters as it is to their ratios. The linear instability criterion is strongly dependent on the ratio of the lat-

eral friction coefficients and the ratio of the vertical friction coefficients, and only on their ratios. The presence of even minimal internal dissipation quickly prevents linear hurricane-like instability for ambient conditions in regions prone to intense tropical cyclones.

Acknowledgments

References

Betts, Alan K. 1982: Saturation point analysis of moist convective overturning; *J. Atmos. Sci.* **39** (7), 1484–1505. Also appeared in an earlier shortened form as 1982: Convective overturning and the saturation point, in E. M. Agee and T. Asai, eds., *Cloud Dynamics*, Reidel.

Black, Peter G. and Richard A. Anthes 1974: On the asymmetric structure of the tropical cyclone outflow layer; *J. Atmos. Sci.* **28** (8), 1348–1366.

Bretherton, Christopher S. 1987: A theory for nonprecipitating moist convection between two parallel plates. Part I: Thermodynamics and "linear" solutions; *J. Atmos. Sci.* **44** (14), 1809–1827.

Bosart, Lance F. and Song C. Lin 1984: A diagnostic analysis of the President's day storm of February 1979; *Mon. Wea. Rev.* **112** (11), 2148–2177.

Burpee, Robert W. 1972: The origin and structure of easterly waves in the lower troposphere of North Africa; *J. Atmos. Sci.* **29** (1), 77–90.

Challa, Malakondayya and Richard L. Pfeffer 1990: Formation of Atlantic hurricanes from cloud clusters and depressions; *J. Atmos. Sci.* **47** (7), 909–927.

Charney, Jule G. and Arnt Eliassen 1964: On the growth of the hurricane depression; *J. Atmos. Sci.* **21** (1), 68–75. This was a slight extension of a similarly titled piece in *Geofis. Int.* **4** (3), 223–230.

Durst, C. S. and R. C. Sutcliffe 1938: The importance of vertical motion in the development of tropical revolving storms; *Quart. J. Roy. Meteor. Soc.* **64** (273), 75–84; with addendum on p. 240.

Emanuel, Kerry A. 1986: An air-sea interaction theory for tropical cyclones. Part I: Steady-state maintenance; *J. Atmos. Sci.* **43** (6), 585–604.

Emanuel, Kerry A., Maurizio Fantini and Alan J. Thorpe 1987: Baroclinic instability in an environment of small stability to slantwise moist convection. Part I: Two-dimensional models; *J. Atmos. Sci.* **44** (12), 1559–1573.

Emanuel, Kerry A. and Richard Rotunno 1989: Polar lows as arctic hurricanes; *Tellus* **41A** (1), 1–17.

Fraedrich, Klaus and John L. McBride 1988: The physical mechanism of CISK and the free-ride balance; *J. Atmos. Sci.* **46** (17), 2642–2648.

Frank, Neil L. and Paul J. Hebert 1974: Atlantic tropical systems of 1973; *Mon. Wea. Rev.* **102** (4), 290–295.

Frank, William M. 1977: Convective fluxes in tropical cyclones; *J. Atmos. Sci.* **34** (10), 1554–1568.

Gray, William M. 1975: Tropical cyclone genesis; Dept. Atmospheric Sci. Paper No. 234, Colorado State University, Fort Collins, 121pp.

Gyakum, John R. 1983: On the evolution of the QE II storm. II: Dynamic and thermodynamic structure; *Mon. Wea. Rev.* **111** (6), 1156–1173.

Handel, Mark David 1990: *Tropical Cyclone Intensification From Finite Amplitude Disturbances, or How Hurricanes Hardly Happen*, Sc.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 261pp.

Handel, Mark David 1991: Tropical cyclone intensification. Part II: Growth from finite amplitude disturbances, ms, Phillips Laboratory/GPAS, Hanscom AFB, Massachusetts, December, to be submitted to *J. Atmos. Sci.*

Handel, Mark David 1991: Tropical cyclone intensification. Part III: Development of an eye, ms, Phillips Laboratory/GPAS, Hanscom AFB, Massachusetts, December, to be submitted to *J. Atmos. Sci.*

Haque, S. M. A. 1952: The initiation of cyclonic circulation in a vertically unstable stagnant air mass; *Quart. J. Roy. Meteor. Soc.* **78**, 394–406.

Hawkins, Harry F. and Daryl T. Rubsam 1968a: Hurricane Hilda, 1964. I. Genesis as revealed by satellite photographs, conventional and aircraft data, *Mon. Wea. Rev.* **96** (7), 428–452.

McBride, John L. 1981: Observational analysis of tropical cyclone formation. Part I: Basic description of data sets; *J. Atmos. Sci.* **38** (6), 1117-1131.

Merrill, R. T. 1988: Environmental influences on hurricane intensification; *J. Atmos. Sci.* **45** (11), 1678–1687.

Molinari, John and David Vollaro 1989: External influences on hurricane intensity. Part I: Outflow layer eddy angular momentum fluxes; *J. Atmos. Sci.* **46** (8), 1093–1105.

Ooyama, Katsuyuki 1964: A dynamical model for the study of tropical cyclone development; *Geofis. Int.* **4** (3), 187–198.

Ooyama, Katsuyuki 1969: Numerical simulation of the life cycle of tropical cyclones; *J. Atmos. Sci.* **26** (1), 3–40.

Orszag, Steven A. and Anthony T. Patera 1983: Secondary instability of wall bounded shear flow; *J. Fluid Mech.* **128**, 347–385.

Palmén, Erik 1948: On the formation and structure of tropical hurricanes; *Geophysica* **3** (1), 26–38.

Palmén, E. 1956: Formation and development of tropical cyclones; *Proc. Tropical Cyclone Symp., Brisbane, Australia*, 213–231.

Rasmussen, E. 1979: The polar low as an extratropical CISK disturbance; *Quart. J. Roy. Meteor. Soc.* **105**, 531–549.

Rasmussen, Erik and Magne Lystad 1987: The Norwegian polar lows project: A summary of the International Conference on Polar Lows; 20–23 May 1986, Oslo, Norway; *Bull. Amer. Met. Soc.* **68** (7), 801–816.

Reed, Richard J. and Mark D. Albright 1986: A case study of explosive cyclogenesis in the eastern Pacific; *Mon. Wea. Rev.* **114** (12), 2297–2319.

Reed, Richard J. and E. E. Recker 1971: Structure and properties of synoptic-scale wave disturbances in the equatorial western Pacific; *J. Atmos. Sci.* **28** (7), 1117–1133.

Schubert, Wayne H., James J. Hack, Pedro L. Silva Dias, and Scott R. Fulton 1980: Geostrophic adjustment in an axisymmetric vortex; *J. Atmos. Sci.* **37** (7), 1464–1484.

Schubert, Wayne H. and James J. Hack 1982: Intertial instability and tropical cyclone development; *J. Atmos. Sci.* **39** (8), 1687–1697.

Xu, Kuanman 1987: Vertical structure and the convective characteristics of the tropical atmosphere; M.S. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 121pp.

Xu, Kuan-man and Kerry A. Emanuel 1989: Is the tropical atmosphere conditionally unstable?; *Mon. Wea. Rev.* **117** (7), 1471–1479.

Properties of Linear Stationary Solutions

	Parameters				Eigenvalues		Ratio of Maxima	
	M	K	B	r_1	$\eta_s^{(0)}$	r_0	R_2	R_3
1)	0.1	0.1	0.002	10.0	1.474	1.969	2.0	0.4
2)	0.1	0.2	0.002	8.0	1.627	2.328	2.2	0.5
3)	0.1	1.0	0.002	10.0	2.284	3.520	4.9	1.1
4)	0.2	0.1	0.005	8.0	1.720	1.675	1.4	0.7
5)	0.2	0.2	0.005	10.0	1.932	2.002	1.9	0.8
6)	0.2	0.5	0.005	10.0	2.370	2.526	2.9	1.3
7)	0.5	0.2	0.005	6.0	2.644	1.633	1.2	1.5
8)	0.5	0.2	0.005	10.0	2.641	1.638	1.4	1.6
9)	0.5	0.2	0.005	12.0	2.642	1.637	1.4	1.6
10)	0.5	0.2	0.020	6.0	2.680	1.599	1.2	1.5
*11)	0.5	0.2	0.050	6.0	2.745	1.542	1.1	1.5
12)	0.5	0.2	0.050	8.0	2.753	1.533	1.2	1.6
13)	0.5	0.2	0.050	10.0	2.756	1.5$_{2,3}$	1.2	1.6
14)	0.5	0.5	0.005	6.0	3.367	2.054	1.9	2.1
15)	0.5	0.5	0.005	8.0	3.357	2.074	2.1	2.2
16)	0.5	0.5	0.005	10.0	3.358	2.073	2.3	2.3
17)	0.5	0.5	0.005	12.0	3.359	2.071	2.3	2.3
18)	0.5	0.5	0.005	14.0	3.360	2.070	2.3	2.3
19)	0.5	1.0	0.005	8.0	4.252	2.468	3.0	3.0
20)	1.0	0.1	0.005	6.0	3.156	1.157	0.8	2.1
21)	1.0	0.5	0.005	8.0	4.749	1.757	1.7	3.6

Table 1: Properties of stationary solutions. Eigenvalues of $\eta^{(0)}$ and r_0 for various sets of parameter values. Also included are the ratios $R_j \equiv |\max(v_j)/\max(v_1)|$. The asterisk indicates the linear solution shown in figure 4 and used in most of the nonlinear calculations of Part II.

Figure 1: Model structure and cumulus parameterization scheme. The region with $0 \leq r \leq r_0$ is the ascent region, while where $r_0 < r \leq r_1$ is the descent region. The boundary layer has fixed height, h_0. The heights of the upper two layers may vary. For every unit mass of air with equivalent potential temperature equal to χ_0 forced out of the boundary layer, $\eta - 1$ units of air with equivalent potential temperature equal to χ_1 are entrained. When the combined η units of mass reach the upper layer they are assumed to be neutrally buoyant with saturated equivalent potential temperature equal to χ_2, which is the same as is the environment. The system is cylindrically symmetric. See text for remaining variable definitions.

Figure 2: Minimum value of the forcing parameter, $\eta^{(0)}$, for the existence of real modes as a function of the vertical and lateral diffusion ratios, M and K, with the cooling parameter set at B = 0.01. This is the requirement for all m^2 to be real and at least one of them positive.

Figure 3: Nondimensional growth rate as a function of squared wavenumber for several values of the forcing parameter, $\eta^{(0)}$. These curves apply to functions consistent with the time dependent dispersion relation, not solutions to the full boundary value problem. The lowest curve, for $\eta^{(0)} = 2.74$, has a maximum growth rate of 0. See text for additional parameter values.

Figure 4: A solution to the linear stationary problem with M = 0.5, K = 0.2, B = 0.05, r_1 = 6.0. The velocities $v_1^{(1)}$ and $v_2^{(1)}$ are the upper and lower layer tangential velocities, respectively, and $v_3^{(1)}$ is proportional to the lower layer inflow velocity. The $v_j^{(1)}$ are all equal to zero at the origin and can be distinguished according to the key in the upper right corner. The line for $\frac{drv_1^{(1)}}{r\,dr}$ has a positive value at the origin and is proportional to the Ekman pumping. There are dashed lines near the upper and lower borders noting the location of r_0.

Figure 5: Growth rate as a function of forcing parameter for the disturbance shown in figure 4. The growth rate is scaled by the Coriolis parameter.

Figure 6: Variations in the neutral forcing parameter, $\eta^{(0)}$, and the updraft extent, r_0, for linear stationary solutions as a function of vertical diffusion ratio, M. Curves are presented for various values of the lateral diffusion ratio, K.

Figure 7: Variations in relative anticylone strength, R_2, and relative inflow strength, R_3, for linear stationary solutions as a function of vertical diffusion ratio, M. Curves are presented for various values of the lateral diffusion ratio, K.

FIGURE 1

FIGURE 2·

194

FIGURE 3

FIGURE 4

196
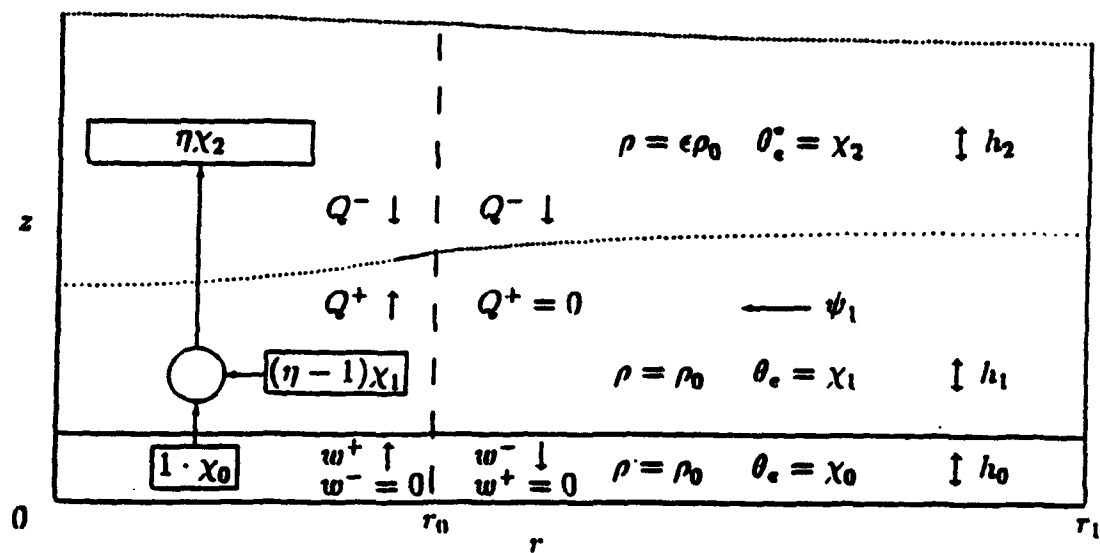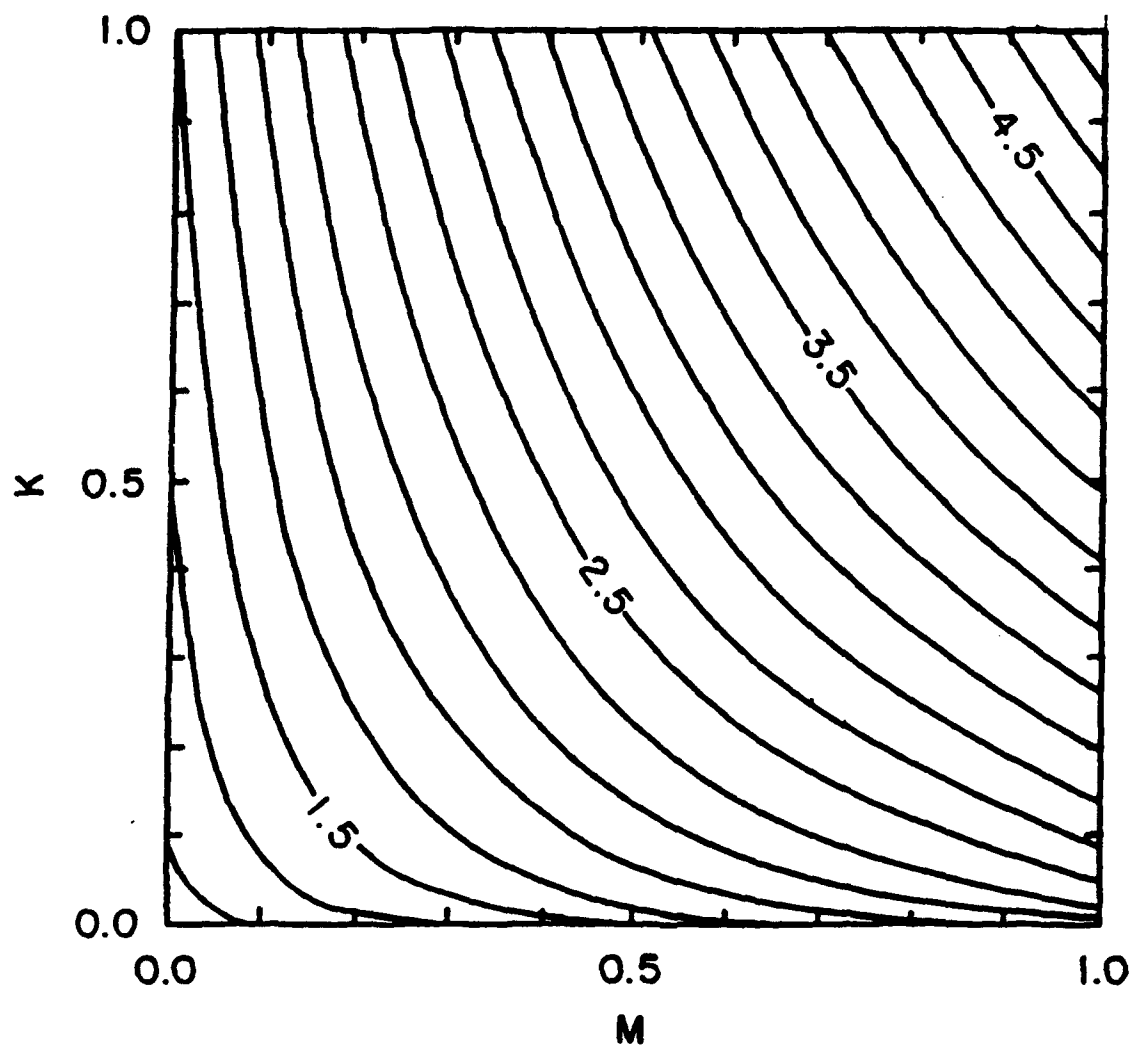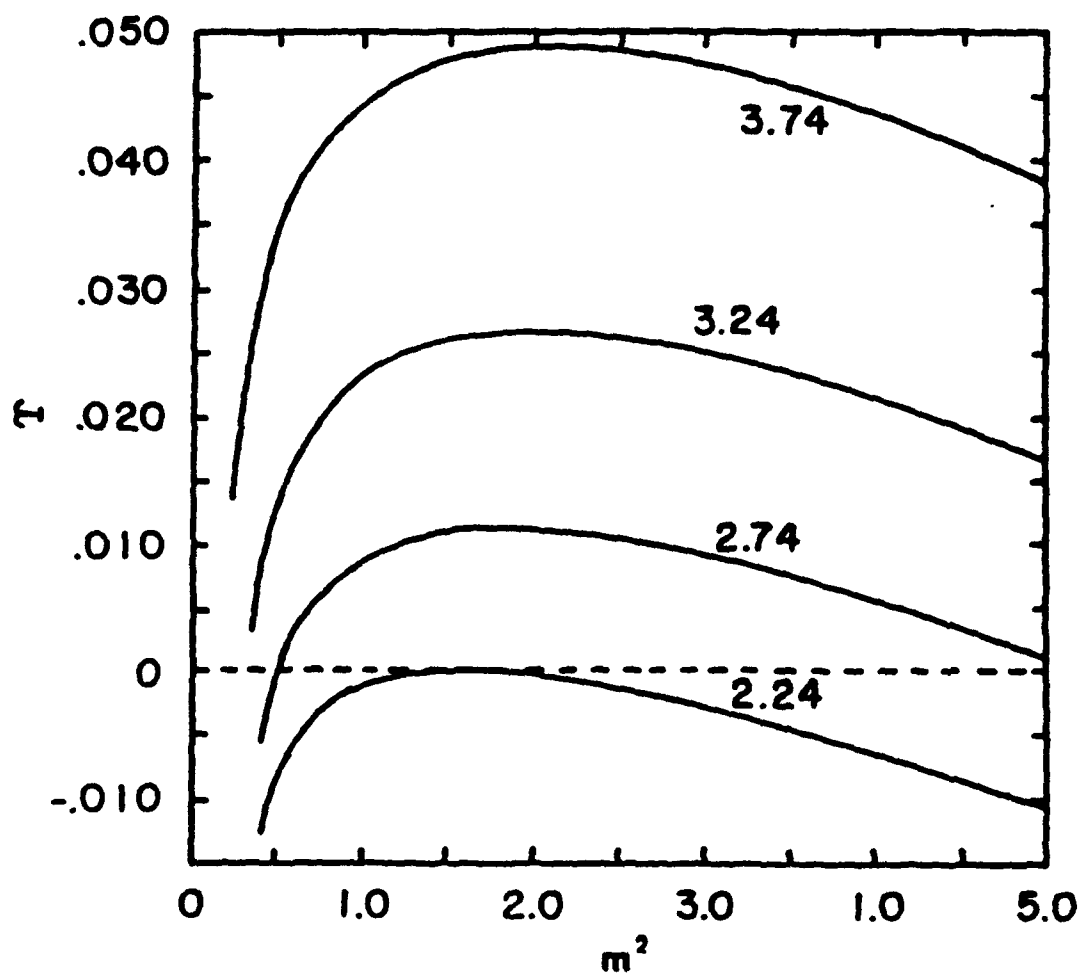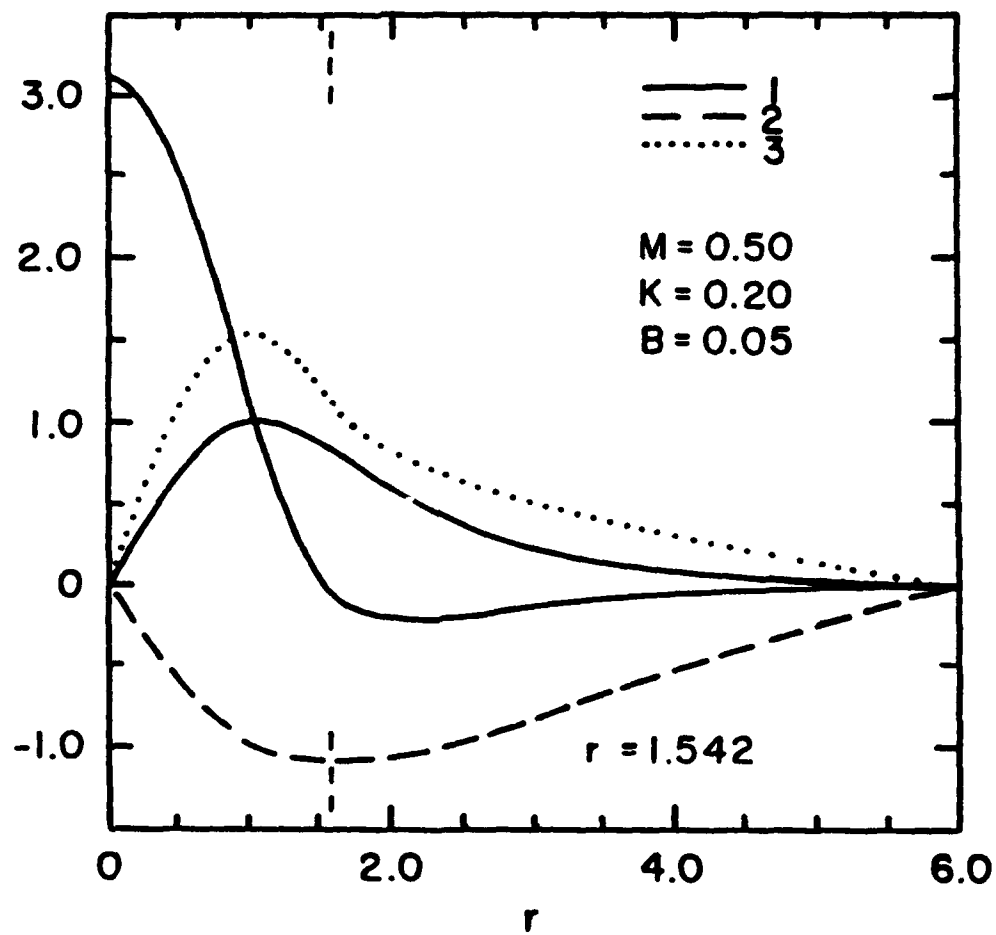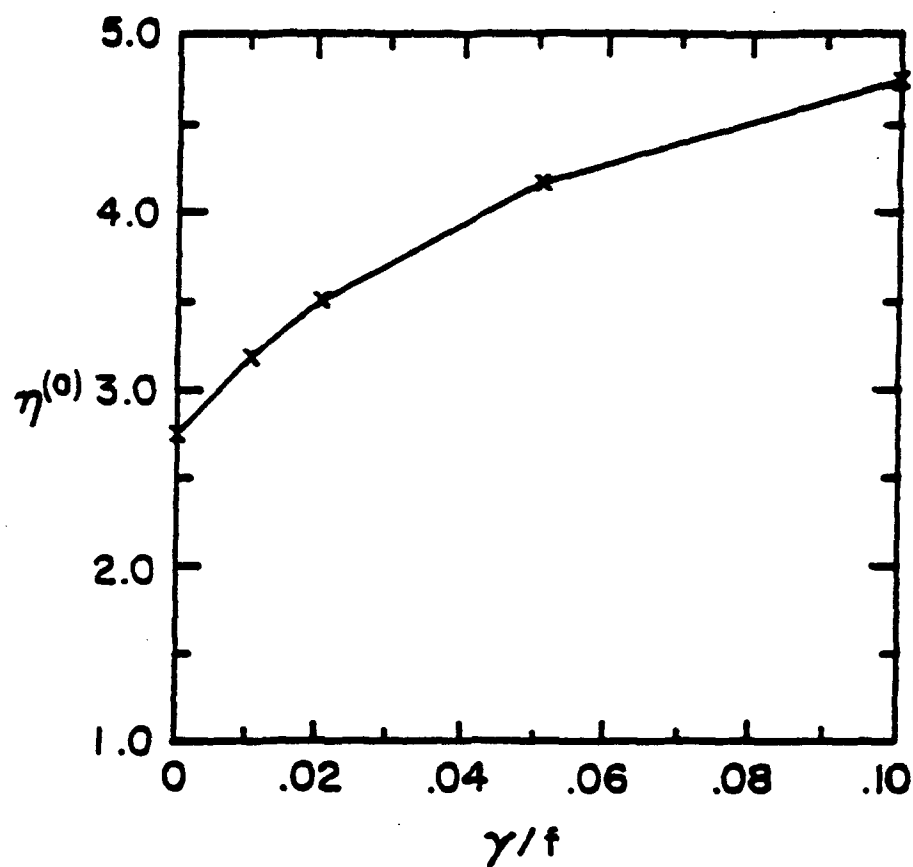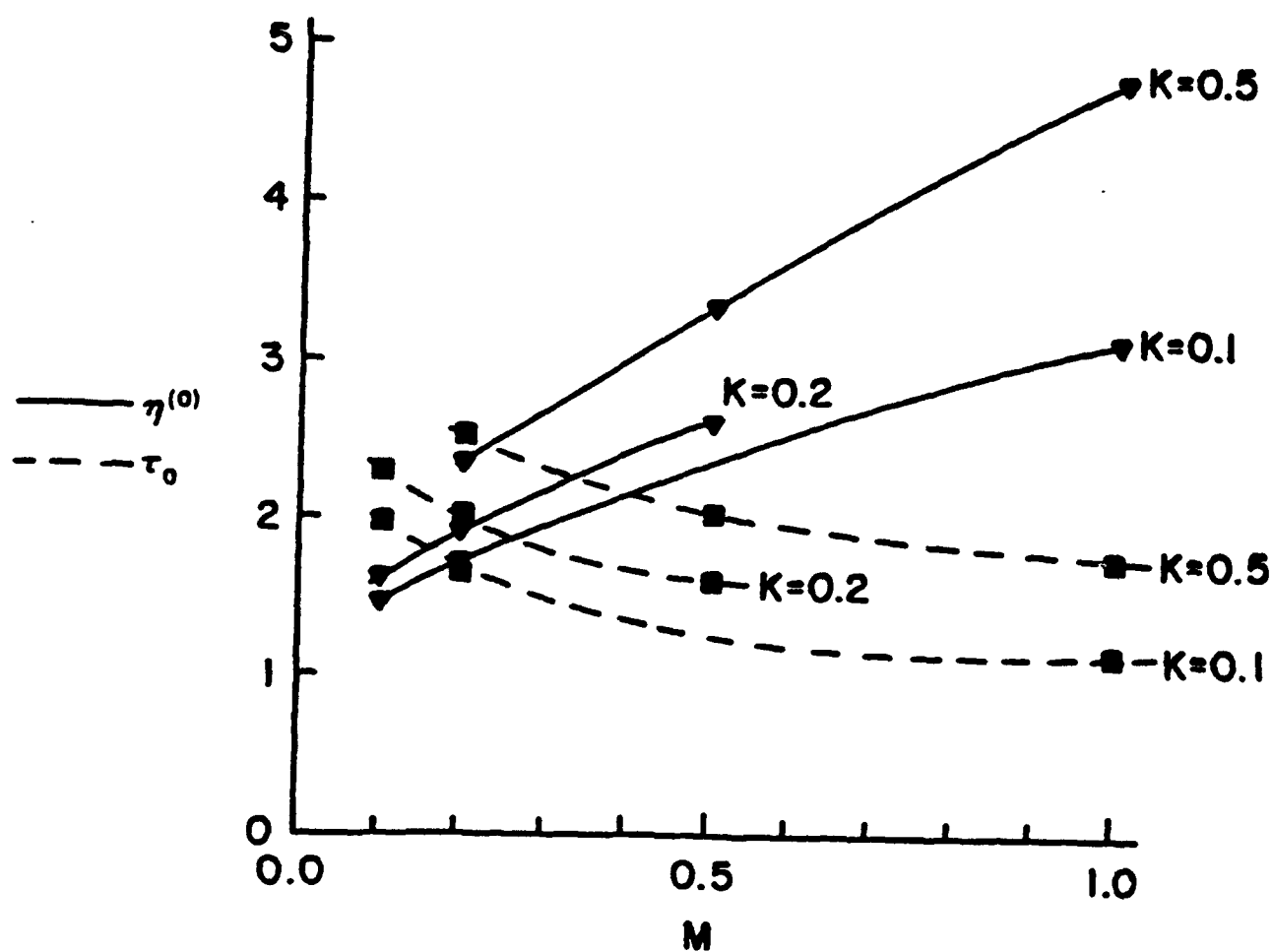
FIGURE 5
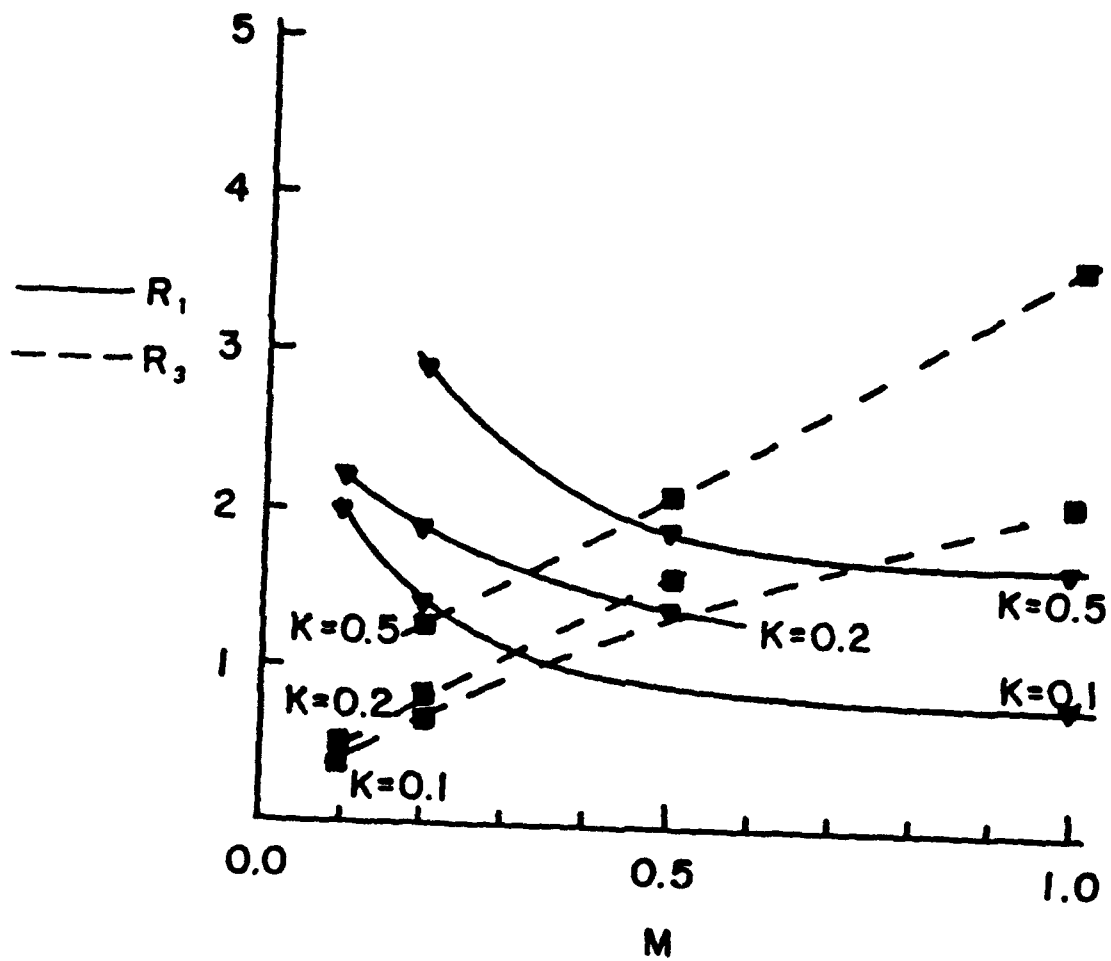
FIGURE 6
197

FIGURE 7

TROPICAL CYCLONE INTENSIFICATION
PART II: GROWTH FROM FINITE AMPLITUDE
DISTURBANCES

Mark David Handel

SCEEE Geophysics Scholars Program

Phillips Laboratory /GPAS

Hanscom AFB, Massachusetts 01731

U.S.A.

December 17, 1991

ABSTRACT

Linear theories of tropical cyclone development suffer from a conspicuous deficiency. If linear instability from small perturbations were possible, then the build-up of potential energy to allow very intense disturbances would be prevented by more frequent weaker disturbances. There appears to be an amplitude threshold for whether or not disturbances intensify. An analytic theory is provided to examine this threshold. In Part I, a linearized viscous CISK (Conditional Instability of the Second Kind) model was examined and it was shown that growth of hurricane-like disturbances from infinitesimal perturbations was unlikely. Here, in Part II, we examine the evolution of finite amplitude disturbances.

We find that finite amplitude disturbances can grow under conditions that are linearly stable. The instability that is found is essentially a subcritical finite amplitude version of CISK. Several nonlinear terms contribute: the faster than linear increase in boundary layer convergence, due to faster than linear increase of the surface drag; the advection of disturbance relative vorticity by the disturbance; and the increase in boundary layer entropy due to elevated surface fluxes of water vapor and sensible heat. For disturbances that exceed the threshold amplitude, growth results that is faster than exponential in time.

1. Introduction

In an examination of west Atlantic hurricane formation, Riehl (1948) found that *"deepening began, without exception, in pre-existing perturbations.... There was no evidence of spontaneous formation due to convection over an overheated tropical ocean. [Emphasis in original.]"* Many tropical cyclones can be traced back to disturbances that have existed for days and travelled thousands of kilometers without significant intensification. In Part I of this study (Handel 1991) we showed that linear growth of tropical cyclones from small disturbances was unlikely even in warm regions of the tropics during summertime. Here, in Part II, we show that if a finite amplitude disturbance with a sufficiently strong surface cyclone travels into, or forms in, a region with some convective available potential energy (CAPE), rapid intensification is possible.

The instability found in this work is essentially a finite amplitude version of Conditional Instability of the Second Kind (CISK) as presented in Ooyama 1969. For a given amount of buoyancy, of parcels in the lower boundary layer with respect to air aloft, and a given stability in the middle troposphere, a threshold amplitude for intensification is found. Disturbances of greater amplitude will continue to grow, while lesser ones will decay.

There are no earlier analytic theories that exhibit a threshold for intensification. The major competition for this theory can be found in the numerical study by Emanuel (1989) on what he has more recently referred to as Wind Induced Surface Heat Exchange (WISHE) theory (Emanuel 1991).

1.1. Nonlinear Growth

Nonlinear theory is used here, not because the amplitudes are so large that linear theory is no longer quantitatively correct (though that is the case) but rather, because

linear theory is fundamentally incapable of exhibiting finite amplitude threshold behavior. Even some of those workers (e.g. Charney and Eliassen 1964, Ooyama 1964) who applied linear theory to the problem of tropical cyclone intensification realized the need for finite amplitude theories and the particular inadequecies of assuming growth from infinitesimal perturbations. The requirement that a rather intense disturbance, within some limited range of configurations, is needed for intensification has appeared in some numerical models (e.g. Rotunno and Emanuel 1987, DeMaria and Pickle 1988, Tuleya 1988, Emanuel 1989). Numerical models that produce hurricane like disturbances with tiny perturbations are linearly unstable and unrealistic, at least with respect to their initial conditions (e.g. Anthes 1972, Yamasaki 1977).

Kleinschmidt (1951) thought that rapid intensification began when the anticyclone aloft became inertially unstable, which implies finite amplitude behavior. A trigger point based on the magnitude of the nonlinear momentum advection terms was proposed by Shapiro (1977). Emanuel (1989), in contradistinction to the work presented here, has recently suggested that Ekman pumping must perform work against a stable gradient, presenting a threshold to be overcome, and that development only occurs if the vortex is strong enough and lasts long enough to increase the moist entropy of the lower troposphere. The qualitative hypothesis of Riehl (1948, 1954) requires the interaction of two independent finite amplitude disturbances, with an upper level trough triggering the further development of a weak low level cyclonic circulation. Elaborations of this "outside influence" triggering hypothesis have been made by Molinari and Vollaro (1989); Challa and Pfeffer (1990); Reilly and Emanuel (1991); and Montgomery and Farrel (1991); all who consider forcing from upper tropospheric disturbances.

In this work we show three nonlinear processes that facilitate tropical cyclone intensification. Since boundary layer friction increases faster than linearly, convection initiated by Ekman pumping increases faster than linearly, causing a nonlinear CISK

development. Also, the nonlinear momentum advection in the lower cyclonic circulation, suggested by Shapiro (1977), provides strong assistance for intensification, and further that a similar process assists in the upper anticyclone. Lastly, the increased moisture flux at the sea surface from increased winds also drives intensification.

1.2. Solution Procedures

The theory presented here is based on the CISK type model of Ooyama (1969). We will examine Ooyama's general system, with the inclusion of dissipation, up to second order in an amplitude expansion. In Part I, the basic equations were discussed in detail and the linearized system was solved. Here in Part II, we continue with a weakly nonlinear analysis to find conditions that allow for subcritical intensification of finite amplitude disturbances.

Expansion of the basic equations to second order results in a system that is linear and inhomogeneous. The associated homogeneous system is identical to the first order linear system, and the inhomogeneous terms are functions of the first order solutions. Inhomogeneous two point boundary problems do not in general have solutions. Ince (1926) provided a detailed theoretical analysis of the algebraic properties of such differential systems. The solvability condition for the two point inhomogeneous case provides a relation between the amplitude of the first order solutions and the forcing parameter (Malkus and Veronis 1958). The forcing parameter is based on the relative stability of the layers, which is closely tied to the convective available potential energy (CAPE). Steady state balance requires that increases in amplitude, though only in the cyclonic sense, be matched by *decreases* in the forcing parameter. With sufficient amplitude for a disturbance, steady state can be maintained with the forcing parameter less than the critical value for linear instability. Existence of a stationary solution is no guarantee of its stability. In fact, it would be a great disappointment if the stationary solution was found to be stable.

Next, equations are derived for perturbations to the steady solutions. This also generates a bounded inhomogeneous linear system with a solvability condition. Since the relation between stationary amplitude and the forcing parameter has already been determined, this second solvability condition then determines the growth rate for a perturbation to the stationary solution (Schlüter *et al.* 1965). Demonstration that the steady solutions are unstable, with a forcing parameter less than critical for linear instability, and unstable in a manner that resembles an intensifying vortex, completes the calculation.

2. Review of Basic Equations

The basic equations used closely follow those of Ooyama 1969. They are also covered at length in Part I. The system is set in cylindrical coordinates (r, z) on a constant rotation f-plane. There are two main layers (lower 1, upper 2) with a boundary layer (layer 0) of fixed height below. All the layers have constant density, with the densities of the boundary layer and lower layer equal. The equations for continuity are

$$0 = \frac{\partial \psi_0}{r \, \partial r} - w \tag{2.1}$$

$$\frac{\partial h_1}{\partial t} = \frac{\partial \psi_1}{r \, \partial r} - Q + w \tag{2.2}$$

$$\epsilon \frac{\partial h_2}{\partial t} = \frac{\partial \psi_2}{r \, \partial r} + Q \tag{2.3}$$

$w \equiv w^+ - w^- \equiv$ vertical velocity out of boundary layer
$\quad [w^+ \text{ (upward)} \geq 0, \; w^- \text{ (downward)} \geq 0]$;

$Q \equiv Q^+ - Q^- \equiv$ interlayer mass flux between layers 1 and 2
$\quad [Q^+ \text{ (upward)} \geq 0, \; Q^- \text{ (downward)} \geq 0]$.

The h_j represent the layer thicknesses and the ψ_j are the *inward* mass fluxes, for layers $j = 0, 1, 2$. Each of the layers has a constant density. For simplicity, the

densities of the boundary and lower layers are assumed equal. The ratio of the density in the upper layer to the density of the boundary and lower layers is ϵ.

After some manipulation (see Part I), the momentum equations can be expressed as

$$h_1 \frac{\partial v_1 r}{\partial t} = \zeta_1 \psi_1 + (Q^- + \mu)(v_2 - v_1)r + \frac{\partial \Lambda_1}{r \, \partial r} \tag{2.4}$$

$$\epsilon h_2 \frac{\partial v_2 r}{\partial t} = \zeta_2 \psi_2 + (Q^+ + \mu)(v_1 - v_2)r + \frac{\partial \Lambda_2}{r \, \partial r} . \tag{2.5}$$

The v_j are the tangential velocities and the $\zeta_j \equiv f + \frac{\partial r v_j}{r \partial r}$ are the total vorticity. Vertical diffusion between the two main layers is included with a linear coefficient μ and the lateral diffusion is Fickian with respect to solid body rotation, so $\Lambda_j = k_j r^3 \frac{\partial}{\partial r}\left(\frac{v_j}{r}\right)$, where the k_j are layer integrated diffusion coefficients.

The pressure field will be assumed to be hydrostatic. The lowest order balance is the gradient wind:

$$\left(f + \frac{v_j}{r}\right) v_j = \frac{\partial \phi_j}{\partial r} , \tag{2.6}$$

where the ϕ_j are the geopotential perturbations and f is the Coriolis parameter. With this balance, the hydrostatic approximation, and the densities of the boundary and lower layers equal, then $v_0 = v_1$ (see Part I for more detail).

By combining the continuity and momentum equations for the constant depth boundary layer, along with a surface stress parameterization, $\tau_s = \rho_0(k_s v_0 + C_D|v_0|v_0)$, we obtain an expression for the mechanical pumping out of the boundary layer:

$$w = \frac{\partial}{r \, \partial r}\left[\frac{k_s r v_1 + C_D r |v_1| v_1 + h_0 \frac{\partial r v_1}{\partial t}}{\left(f + \frac{\partial r v_1}{r \partial r}\right)}\right] . \tag{2.7}$$

The surface stress has a linear term proportional to k_s and a quadratic drag proportional to C_D, while ρ_0 is the boundary layer density. This will be referred to

as the Ekman pumping, though technically the time dependent term should not be included under that rubric. Comparison of the linear and nonlinear terms will be made in Part III (Handel 1992).

It is assumed that cumulus convection only occurs where air is mechanically forced out of the boundary layer. The primary effect of convection is to drive entrainment in response to heating. The vertical flow from layer 1 to layer 2 is set proportional to the positive Ekman pumping and an entrainment parameter η:

$$Q^+ = \eta w \,, \tag{2.8}$$

with

$$\eta = \begin{cases} 1 + \dfrac{\chi_0 - \chi_2}{\chi_2 - \chi_1} \,, & w > 0 \quad (w^+ > 0) \\ 0 \,, & w \leq 0 \quad (w^+ = 0) \end{cases} \tag{2.9}$$

where $\chi_0 \equiv \theta_e$ of the boundary layer, $\chi_1 \equiv \theta_e$ of the lower layer, $\chi_2 \equiv \theta_e^*$ of the upper layer, θ_e is the equivalent potential temperature, and θ_e^* is the saturation equivalent potential temperature. The parameter η is also the forcing parameter for the system.

Newtonian cooling is included throughout the domain. By use of the assumptions that most of the temperature perturbation is in the upper layer and that layer thickness between pressure surfaces is proportional to temperature, along with the gradient wind balance, yields an expression for the radial derivative of the downward motion driven by radiative cooling:

$$\frac{\partial Q^-}{\partial r} = b \left[\left(f + \frac{v_2}{r} \right) v_2 - \left(f + \frac{v_1}{r} \right) v_1 \right] \,, \tag{2.10}$$

where b is a coefficient proportional to the cooling rate. The details of these parameterizations of the diabatic processes are discussed in detail in Part I.

For the calculations at second order, we will need the first order perturbations

to the entropy of each layer. In the boundary layer

$$\frac{\partial \chi_0}{\partial t} - \frac{\psi_0}{h_0}\frac{\partial \chi_0}{r\,\partial r} + \frac{w^-}{h_0}(\chi_0 - \chi_1) =$$

$$\frac{C_\varepsilon}{h_0}|v_1|(\chi_s - \chi_0) + \frac{\partial}{r\,\partial r}\left(k_\chi r \frac{\partial \chi_0}{\partial r}\right) , \qquad (2.11)$$

where C_ε is a bulk exchange coefficient and k_χ is a diffusion coefficient. The equivalent potential temperature of a parcel saturated at the sea surface temperature is designated χ_s. The equivalent potential temperature of the lower layer is assumed to be constant, though as will be discussed later this may not have been wise. In the upper layer the perturbations will be approximated as

$$\chi_2 = \chi_2^{(0)} + \Xi(\theta_m - \theta_m^{(0)}) , \qquad (2.12)$$

where θ_m is a mean column potential temperature and $\theta_m^{(0)}$ is the mean column potential temperature of the rest state. This assumes that changes in the saturation equivalent potential temperature are Ξ times the perturbations of the potential temperature in the upper troposphere. We will set $\Xi = 1.2$. The mean potential temperature perturbation can be obtained by radially integrating an expression proportional to the right hand side of eq. (2.10).

3. Finite Amplitude Disturbances

3.1. Expansions and the Linear System

The fields will be expressed with an amplitude expansion similar in purpose to that used by Malkus and Veronis (1958), or by Schlüter et al. (1965), in their weakly nonlinear analyses. In each of the layers, the azimuthal velocity is expanded as

$$v_j = a v_j^{(1)} + a^2 v_j^{(2)} + \cdots \qquad (3.1)$$

The amplitude, a, is a Rossby number measured as $a \equiv V/(fL)$, where V is the magnitude of the velocity perturbation and L is the length scale (not yet determined). All other variables will be expanded similarly. A superscript of (0) will be used for values applying to the rest state.

The time dependent expansion will be an infinitesimal linear perturbation, with amplitude $\tilde{a}$, about a small amplitude stationary state, with amplitude $\bar{a}$ to be determined, such that $a = \bar{a} + \tilde{a}$ though $\tilde{a} \ll \bar{a}$. In particular, the expansions require that the spatial functions $\bar{v}_1^{(1)}$ of the steady first order solutions and the infinitesimal time dependent perturbations $\tilde{v}_1^{(1)}$ are identical, since at first order stationary and slowly varying solutions are both solutions to the identical linear system. So, $\bar{v}_1^{(1)} = \bar{a} v_1^{(1)}$ and $\tilde{v}_1^{(1)} = \tilde{a} v_1^{(1)}$, where $v_1^{(1)}$ is a solution to the linear stationary problem of normalized amplitude.

Time dependent perturbations will have an assumed time dependence of $e^{\gamma t}$. The growth rate is expanded as

$$\gamma = \gamma^{(0)} + (\bar{a} + \tilde{a})\gamma^{(1)} + a^2\gamma^{(2)} + \cdots . \tag{3.2}$$

For solutions that are stationary, $\gamma = 0$. A solution for which $\gamma^{(0)} = 0$, but $\gamma^{(1)} \neq 0$ will be referred to as *slowly varying*.

To convert the linear problem into a vector form, two new variables are introduced: $v_3 \equiv (f/k_s)\psi_1/r$, $v_4 \equiv (f/k_s)\psi_2/r$. For stationary and slowly varying solutions (the only ones with which we are concerned here), imposition of the boundary conditions that the lateral flow vanishes at the center and outer limit of the disturbance along with the sum of the continuity equations (2.1)–(2.3), requires $v_1^{(1)} + v_2^{(1)} + v_4^{(1)} = 0$. This allows reduction of the linear system to one for the vector $\mathbf{v} = (v_1, v_2, v_3)$:

$$r\frac{d}{dr}\left(k_2^{(0)}B\frac{dr\mathbf{v}^{(1)}}{r\,dr}\right) + k_s A r \mathbf{v}^{(1)} = 0 . \tag{3.3}$$

Appendix A provides the matrices, A and B, of the linear system. Matrix A is constant and matrix B is piecewise constant with jump discontinuities at radii dividing ascent regions from descent regions. The boundary conditions for this viscous system are that the flow vanish at the center and at the outer boundary of the domain.

We will only be concerned with solutions having a single ascent region surrounding the origin and a single descent region outside of that. At the dividing radius, r_0, v and its first radial derivative must be continuous. This radius is an eigenvalue of the linear system. Figure 1 shows an example of a solution to the linear system as found in Part I. The eigenvalue of the forcing parameter for an actual solution is designated $\eta_s^{(0)}$.

The most complicated expansion involves the forcing parameter, η. Unlike in most stability problems, the forcing parameter can be modified by the *first* order solutions through changes in the χ_j. The part of η that may vary with the $\chi_j^{(1)}$ is designated, $\eta_r^{(1)}$. It is a function of the motion and may depend on r. There is still a need to expand the constant part of η, as is done in most weakly finite amplitude problems. The required expansion part, η_c, has no spatial dependence and is determined with a solvability condition, as will be explained below. To second order then,

$$\eta = \eta^{(0)} + (\bar{a} + \tilde{a})\eta_c^{(1)} + (\bar{a} + \tilde{a})\eta_r^{(1)}(r) + a^2\eta^{(2)} + \cdots . \tag{3.4}$$

Expansion of $\eta^{(2)}$ is not performed since this term will not be needed. The solvability condition, to be derived in section 3.2, will relate the steady state amplitude to the forcing. For $\bar{a}$ positive, if $\eta_c^{(1)} < 0$ then a cyclonic subcritical instability may be possible.[1] By itself, $\eta_c^{(\cdot)} < 0$ only indicates the possiblity of a subcritical stationary

---

[1] Usually if the stability of a system is altered at second order, any resulting bifurcation is referred to as *transcritical.* This can be a subset of supercritical or subcritical instabilites (see Bergé *et al.* 1986, p. 272ff).

solution; it is still necessary to determine the stability of that solution. The expansion for $\eta_r^{(1)}$ is based on a Taylor series expansion of eq. (2.9) after performing an amplitude expansion of the χ_j:

$$\eta_r^{(1)} = \frac{\chi_0^{(1)} - \eta^{(0)}\chi_2^{(1)}}{\chi_2^{(0)} - \chi_1^{(0)}} . \tag{3.5}$$

Boundary layer increases in temperature or moisture causing increases in $\chi_0^{(1)}$ help drive intensification. Warming aloft that increases $\chi_2^{(1)}$ stabilizes the system. The expressions for $\chi_0^{(1)}$, $\chi_2^{(1)}$, and other needed expansions for the second order problem can be found in appendix B.1.

3.2. Amplitude of the Stationary Solutions

The linear solution we have found has, as yet, no determined amplitude. The stationary system can be expanded to second order, the results of which can be obtained from appendix B.2. The solvability condition on the resulting inhomogeneous system will provide a relationship between the forcing parameter and the stationary solution amplitude. It is already assumed from eq. (3.3) that $r\frac{d}{dr}\left(k_2^{(0)}B\frac{dr\bar{v}^{(1)}}{r\,dr}\right)+k_s rA\bar{v}^{(1)} = 0$, where $\bar{v}^{(1)} = \bar{a}v^{(1)}$.

The vector $\bar{v}$ at second order will be expanded as $\bar{v}^{(2)} = \bar{a}^2\left[v^{(2h)} + v^{(2)}\right]$, where $v^{(2h)}$ is the solution to the associated homogeneous equation at second order and $v^{(2)}$ is a particular solution to the inhomogenous problem that is orthogonal to the solutions to the homogeneous problem. The trick of Malkus and Veronis (1958) was to realize that since the spatial structure of $v^{(2h)}$ is identical to $\bar{v}^{(1)}$, it should be absorbed into the coefficient $\bar{a}$ of $\bar{v}^{(1)}$ and then $v^{(2h)}$ taken to be zero everywhere. After reducing the four inhomogenous second order equations to three and transforming the system to self-adjoint form, we can express the result as

$$r\frac{d}{dr}\left(k_2^{(0)}B\frac{dr\bar{v}^{(2)}}{r\,dr}k_s\right) + Ar\bar{v}^{(2)} = rT\mathcal{M}^{(2)} , \tag{3.6}$$

where matrix T is given in eq. (A.4) and $\mathcal{M}^{(2)}$ is the vector of the inhomogenous terms determined in appendix B.2.

The solvability condition for the inhomogeneous system with homogeneous boundary conditions over the domain $[0, r_1]$ is then

$$\int_0^{r_1} r\mathbf{v}^{(1)'T} T\mathcal{M}^{(2)}\, dr = 0 \,.$$

where $\mathbf{v}^{(1)'}$ is a solution to the homogeneous problem with amplitude a' and the superscript T indicates a transpose (see Handel 1990 for a proof of the solvability condition in vector form). This can be expressed as

$$\eta_c^{(1)}\Omega + \Sigma = 0. \tag{3.7}$$

where Σ and Ω are definite integrals shown in table 2. The integral Ω captures all of the terms that are proportional to $\eta_c^{(1)}$, is quadratic in the $v_j^{(1)}$, and is only integrated over the domain $[0, r_0]$ since $\eta_c^{(1)}$ is zero outside of that. As shown in the first entry of table 2, an integration by parts has already been performed. The integral Σ captures all of the the remaining terms, is cubic in the $v_j^{(1)}$, and most of its terms are integrated over the full domain. This integral has been separated into nine terms in the table for comparison of their relative importance. Entries Σ_6–Σ_9 have survived an integration by parts.

From examination of the integral for Ω, in the first entry of table 2, it is clear that Ω is negative. Then for subcritical finite amplitude stationary states to exit, Σ must also be negative. The integral that comprises Σ is much more complicated and will be discussed in section 3.5, where the numerical evaluation of the integrals is considered.

For Σ and Ω both negative, $\eta_c^{(1)}$ is also negative. This allows subcritical stationary states for $\bar{a} > 0$. Most of what has preceded this only applies for positive $\bar{a}$. One could examine this system with anticyclonic circulations in the lower layer, but

211

the assumptions on the location of the heating would no longer apply. The effective value of η for a given finite amplitude disturbance is defined as $\eta_e \equiv \eta_s^{(0)} + \bar{a}\eta_c^{(1)}$. For an ambient forcing parameter η_a and disturbance requiring a forcing parameter $\eta_s^{(0)}$, there is a critical stationary amplitude, a_{crit}, where $\eta_e = \eta_a$ such that

$$a_{crit} = \frac{\eta_a - \eta_s^{(0)}}{\eta_c^{(1)}} = (\eta_s^{(0)} - \eta_a)\frac{\Omega}{\Sigma} \; . \tag{3.8}$$

3.3. Stability of the Stationary Solutions

Having found a stationary solution, it is now possible to look at time dependent perturbations to that solution. The system is separable for the independent variables r and t. With respect to t the equations are linear and constant coefficient, so we may assume solutions proportional to $e^{\gamma t} = e^{a\gamma^{(1)} t}$. At first order we again have $r\frac{d}{dr}\left(k_2^{(0)}B\frac{dr\tilde{v}^{(1)}}{r\,dr}\right) + k_s Ar\tilde{v}^{(1)} = 0$. The slowly varying system can now be expanded to second order, the results of which are shown in appendix B.2.

The vector $\tilde{v}$ at second order will be expanded as $\tilde{v}^{(2)} = 2\bar{a}\bar{a}\left[v^{(2h)} + v^{(2)}\right]$, where $v^{(2h)}$ is the solution to the associated homogeneous equation at second order and $v^{(2)}$ is a particular solution to the inhomogenous problem which is orthogonal to the solutions of the homogeneous problem. As with the stationary case, the spatial structure of $v^{(2h)}$ is identical to $\tilde{v}^{(1)}$. By a renormalization similar to that of the last section, $v^{(2h)}$ is set equal to zero everywhere. After reducing the four inhomogenous second order equations to three and transforming the system to self-adjoint form, we can express the result as

$$r\frac{d}{dr}\left(k_2^{(0)}B\frac{dr\tilde{v}^{(2)}}{r\,dr}\right) + k_s Ar\tilde{v}^{(2)} = rT\mathcal{N}^{(2)} \; ,$$

where $\mathcal{N}^{(2)}$ is the vector of the inhomogenous terms given in appendix B.2.

The solvability condition is

$$\int_0^{r_1} r\mathbf{v}^{(1)'T}\mathbf{T}\mathcal{N}^{(2)}\,dr = 0 \, ,\tag{3.9}$$

which expanded is

$$2\Sigma + \gamma^{(1)}\Gamma + \eta_c^{(1)}\Omega = 0 \, .\tag{3.10}$$

With eq. (3.7) then

$$\gamma^{(1)} = \frac{-\Sigma}{\Gamma} \, .\tag{3.11}$$

The integrals for Γ are given in the third grouping of table 2. Since mostly this integral is dominated by expressions for the kinetic energy and the vortex potential energy, we might expect that Γ is positive.

So if subcritical stationary states exist (implying $\Sigma < 0$) and $\Gamma > 0$, they are unstable. For infinitesimal disturbances off of the stationary state the growth rate is initially

$$\gamma = a_{crit}\gamma^{(1)} \, ,$$

where a_{crit} is determined by eq. (3.8) and the growing disturbance amplitude is that of the perturbation from the stationary state. However, if we continue to take the growth rate as $\gamma = a\gamma^{(1)}$, with a increasing in time, superexponential growth results. This will be discussed further in section 4. As η_a increases approaching $\eta_s^{(0)}$, a_{crit} decreases, decreasing γ at the stationary point. However, for any given disturbance amplitude this just shifts the partitioning of the amplitude between $\bar{a}$ and $\tilde{a}$. Eventually the assumption that $\tilde{a} \ll \bar{a}$ breaks down, but the qualitative results still hold.

3.4. Parameter Values

All parameter values are provided in table 1. The analysis of Xu (1987) was used as a rough guide for determining suitable values of the $\chi_j^{(0)}$. Choosing optimistically to maximize η_a, one could take a boundary layer equivalent potential temperature of $\chi_0^{(0)} \approx 357\,\text{K}$, a lower troposphere equivalent potential temperature of $\chi_1^{(0)} \approx 340\,\text{K}$, and an upper troposphere saturation equivalent potential temperature of $\chi_2^{(0)} \approx 347\,\text{K}$, yielding $\eta_a \approx 2.5$. In choosing $\chi_1^{(0)}$ and $\chi_2^{(0)}$ representative values for the inflow and outflow are desired, rather than the minimum in θ_e for the inflow and the value of θ_e^* at the tropopause for the ouflow. In practice, η_a is usually much smaller than this due to water loading effects and less favorable values of the moist entropies. It is probably even less than 1.5 in most regions of development. For a given value of $\eta_s^{(0)}$ only two of the three $\chi_j^{(0)}$ can be chosen.

The bulk aerodynamic transfer coefficients, C_D and C_E, were both assigned values of order 10^{-3}. The linear surface drag was then chosen to be about $k_s \sim C_D \cdot 5\,\text{m/sec} \sim 5 \cdot 10^{-3}\,\text{m/sec}$. For a given value of vertical diffusion ratio $M = \mu/k_s$, this then fixes the bulk vertical diffusion coefficient μ. Multiplication by the vertical length scale indicates that the values of μ for $M < 1$ imply vertical eddy diffusivities of order $10\,\text{m}^2/\text{sec}$, which are quite small compared to the values chosen in some numerical models.

Once given k_s, the value of $k_2^{(0)} = \epsilon\lambda_2 h_2^{(0)}$ was chosen to keep r_0 of order a couple of hundred kilometers and r_1 a couple of thousand kilometers or less. This implied an upper level eddy viscosity $\lambda_2 \sim 10^4\,\text{m}^2/\text{sec}$. These values are smaller than those chosen for many axisymmetric numerical models. For the values of $K = k_1^{(0)}/k_2^{(0)}$ chosen, λ_1 was smaller still. The much greater asymmetry of the upper level flow compared to the lower level flow provides the justification for having $k_1^{(0)} < k_2^{(0)}$.

With $B = bf^2 k_2^{(0)}/k_s^2$ set; f, $k_2^{(0)}$, and k_s already chosen; the cooling constant b is

also determined and ends up of order 10^{-6} sec/m, which implies a radiative relaxation time of one to two weeks. This cooling rate seems smaller than realistic, but larger values of B require the use of Bessel functions of complex argument, which we wish to avoid.

The parameters k_χ, C_ε, $\chi_s^{(0)}$, and $\chi_1^{(0)}$, appear only in the evaluation of $\chi_0^{(1)}$ in eq. B.2. Since there are only two terms in this equation, this set of parameters is effectively reduced to two. Further, the first term (proportional to $\chi_s^{(0)} - \chi_0^{(0)}$) dominates this formulation since in this model there are no strong downdrafts into the boundary layer. Because this linear (in v_j) expression has no feedback for reducing the entropy difference between the sea surface and the boundary layer, nor is there any boundary layer drying from downdrafts, $\chi_0^{(1)}$ can get so large that $a\chi_0^{(1)}$ can exceed $\chi_s^{(0)} - \chi_0^{(0)}$, which is nonphysical. The lack of any turbulent fluxes of moist entropy from the boundary to lower layer also contributes to the excessive response in the boundary layer. To yield more reasonable values, k_χ was made unrealistically large. Another possible fix might have been to allow for variation in χ_1.

The parameters for the specific heat, c_p; and functions of mean layer pressure $\pi_j = (p_{0j}/p_{00})^{2/7}$, where p_{0j} is the mean pressure of layer j; appear only in the integral for $\chi_2^{(1)}$ in eq. B.4. The specific heat is considered fixed, the π_j vary only slowly with the mean pressure of the layers, f is fixed for a given location, and the factor $\Xi = 1.2$ can only be varied slightly. So, this integral cannot be easily tuned. Since it represents the thermal perturbation needed to maintain gradient wind balance, this is not surprising. Perturbations that are spatially large, intense, or at high latitude, require large temperature anomalies. Plots of both of $\chi_0^{(1)}$ and $\chi_2^{(1)}$ are in figure 2. From these functions and eq. (3.5), one can determine the perturbation to the forcing parameter $\eta_r^{(1)}$, which is plotted in figure 3. To obtain a positive $\eta_r^{(1)}$, it must be that $\chi_0^{(1)} > \eta^{(0)}\chi_2^{(1)}$. One can always tune $\chi_0^{(1)}$ to meet this condition throughout the domain, though for large vortices or at high latitudes this

might require unrealistic increases in the boundary layer entropy.

3.5.   Nonlinear Forcing and Growth Rates

We now examine the terms that make up the integral Σ. To simplify the discussion, it is worth noting that the leading factors in square brackets of Σ_1–Σ_5 are positive everywhere for $B < M/K$. [Use of the linear system can show that $v_1^{(1)}\left(M - \frac{M}{K} - BK\right) + v_2^{(1)}\left(\frac{M}{K} - B\right) + \frac{v_3^{(1)}}{K} = v_1^{(1)}(M - BK) - v_2^{(1)}B - \frac{k}{k_3^{(1)}}\Box^2 v_3^{(1)}$. We know that $v_1^{(1)} > 0$, $v_2^{(1)} < 0$, and from examination of figure 1 that $\Box^2 v_3^{(1)} \equiv \frac{d}{dr}\left(\frac{dr v_3^{(1)}}{r\,dr}\right) < 0$ except by a small amount in a small region just outside of r_0.] Terms that tend to weaken the cyclone are Σ_1, Σ_5, and Σ_8. The first of these represents transport of anticyclonic relative momentum into the lower layer through radiative cooling and weakens the lower level vortex. The two other terms are both tied to second order effects of the forced vertical velocity. Though in developing a tropical cyclone it is desireable to have lower level inflowing air, the accompanying outflow is a net cost. Ideally for development, inflowing air would just vanish, but the outflow is inevitable. Both Σ_5 and Σ_8 are from second order effects that lead to increased outflow aloft and hence increase the strength of the upper level anticyclone. In our example, these terms are all of the same order of magnitude.

There are a number of second order terms that encourage tropical cyclone development. The largest of these in table 2 is Σ_7. This is one of the two nonlinear CISK terms. It is from the increased convection based on the initial stratification and the faster than linear growth of the forced vertical velocity out of the boundary layer. The other nonlinear CISK term, Σ_6, is from increases in the boundary layer entropy from wind enhanced sea surface fluxes (mostly of moisture), partially compensated by an increase in the stable stratification from warming aloft needed to maintain gradient wind balance. The term Σ_6 is partially in the spirit of Emanuel's WISHE

theory, but not quite, since it still relies on frictionally forced vertical velocity to release any of this added potential energy. The competing processes of boundary layer moistening and upper tropospheric warming could be accounted separately, though we will not do so here. The relative magnitudes of these terms, both with respect to each other and with respect to the other second order terms, can be easily changed by adjusting the parameters and should not be viewed as a robust result of this analysis.

The nonlinear momentum advection terms in the lower and upper layers, Σ_2 and Σ_4, both encourage development of the disturbance. In the lower layer it is easy to see that the advection of positive relative vorticity, in addition to the planetary vorticity, helps build the cyclonic flow. The importance of self-advection of vorticity in the inflow layer is consistent with the suggestions of Shapiro (1977). The situation aloft is more subtle. We have already suggested that energy spent in developing the anticyclone hinders the development of the low level cyclone. Aloft, there is negative relative vorticity in the first order solutions, except at large radius. Outward advection of negative relative vorticity reduces the anticyclone. Though at the weak stage examined here the vorticity advection aloft assists development, during the later evolution of the storm, when the cyclone develops to the point where the vorticity of the outflow layer is positive at small radius, the nonlinear advection will change sign in the core region. For solutions requiring a large value of $\eta^{(0)}$, unless the boundary layer entropy is allowed to become very large, the dominant terms at second order are the nonlinear advections of disturbance vorticity. The last term to cover is Σ_3, which is the cumulus momentum mixing term. With negative vertical shear, this tends to reduce the strength of the anticyclone and so help overall disturbance growth.

To maintain a steady state at finite amplitude, with all of these positive influences from the second order terms, the forcing parameter need not be as large as it must be for a steady balance with an infinitesimal disturbance. The term $\eta_c^{(1)}\Omega$ in

eq. (3.7) represents physics similar to that of Σ_6, which is based on changes in the vertical thermodynamic structure due to the first order motion. By fiat, through the solvability condition, it is declared to balance the positive influences of the other second order terms. It therefore provides the quantitative measure of the amount by which the forcing parameter may be reduced to keep steady balance for a given amplitude.

If we do not have an exact balance, but instead have an evolving disturbance, the extra forcing (of either sign) goes into changing the energy of the system. Loosely, the forcing balances the time rate of change of the system energy, or for exponential growth it balances the growth rate times the energy. The integrals that make up Γ can be viewed in some sense as energies of the system. That large positive values of Γ lead to small growth rates, reflects the fact that the greater the energy of a system the more difficult it is to maintain exponential intensification. For rapid growth, small positive values of Γ are required. As in the expansion of Σ, the first coefficient in square brackets for Γ is positive except in a very small region where it is negative by only a small amount. The term proportional to $h_1^{(0)}v_1^{(1)}$ is large and positive, while the one proportional to $\epsilon h_2^{(0)}v_2^{(1)}$ is large and negative. Again, the upper layer persists in its contrary role, by reducing overall Γ through Γ_2. The integrals of Γ_4 and Γ_1 are closely related to the kinetic energies of the boundary and lower layers. The potential energy of tilted isotherms is represented by Γ_3 and Γ_5. The term Γ_6 is related to the time dependent corrections to the forced vertical velocity at the top of the boundary layer.

If one assumes an ambient η_a of say 1.5, then the critical Rossby number is about 0.25. This corresponds to an organized tangential velocity maximum of 2.0 m/sec and a center surface vorticity of $3.9 \cdot 10^{-5}\,\mathrm{sec}^{-1}$, which is nearly $0.8f$ despite the not-large nondimensional amplitude in terms of velocity. The e-folding time is $(4.5 \cdot 10^{-6}\,\mathrm{sec}^{-1})^{-1} \approx 2.5$ days, divided by the Rossby number. With $a_{crit} \approx 0.25$, this

is too long at about 10 days. If one stretches the theory past the stationary point, at a Rossby number of 1, which for these parameters requires a v_1 maximum of only $fL \approx 8\,\text{m/sec}$, the e-folding time of about 2.5 days is quite reasonable for a disturbance of this strength. Even a disturbance requiring a large value of the forcing parameter, in an environment with only small amounts of CAPE, can be destabilized with sufficient amplitude.

3.6. Sensitivity to Parameters

We will now examine how the finite amplitude instability criteria vary with variations of many of the physical parameters. Unless otherwise noted, all of parameters are as specified in table 2. An extensive sensitivity analysis can be found in Handel 1990. In Part I, a sensitivity analysis of the required forcing parameter, $\eta_s^{(0)}$, to changes in the dissipation ratios M, K, and B was presented.

Results from variation of the quadratic bulk diffusion coeffient for the boundary layer, C_D, are shown in table 3. Increases in C_D lead to increases in the magnitude of $\eta_c^{(1)}$, hence decreases in the threshold amplitude for growth, and increases in the growth rate of the finite amplitude disturbances. So, increases in the loss of boundary layer momentum to the surface encourages tropical cyclone development. The only integrals affected by changes in C_D are Σ_5, Σ_7 and Σ_8. The magnitudes of all of these increase with increasing C_D. Though Σ_5 and Σ_8 are both dissipative, these are more than compensated by the additional cumulus heating driven by the additional Ekman pumping. Increases in surface entropy fluxes through increase in C_E can only encourage development and are examined further in Handel 1990.

Results for variations in the Coriolis parameter are shown in table 4. The dimensional vorticity and velocity scales, f and fL, increase linearly with increasing f. Hence, the dimensional perturbation threshold continues to increase with latitude

even though the of magnitude of $\eta_c^{(1)}$ is increasing, because the variation of the latter is weak. This can be seen in the last column of table 4. (If the dimensional magnitude of $-\eta_c^{(1)}/(fL)$ is made smaller, a stronger disturbance is needed to overcome the threshold.) At very low latitudes, where the threshold is quite small dimensionally, though it is easy for a disturbance to start growing, it is initially so weak and the growth rate so small that it would take weeks to become of notice (or hit land or be destroyed by shear). It appears that it is more difficult to initiate development at higher latitude, but if the threshold is exceeded intensification is more rapid.

Proportional variations in the lateral diffusion coefficients $k_1^{(0)}$ and $k_2^{(0)}$ with constant K lead to a confounding change in the length scale of the system through the choice of the nondimensionalization. Results of such variation are shown in table 5. Increases in the diffusion coefficients imply an increase in both the length scale and the velocity scale. Though $\eta_c^{(1)}$ is seen to increase monotonically with increases in $k_1^{(0)}$, seeming to imply a decrease in the threshold amplitude, the dimensional amplitude of the needed disturbance is fairly insensitive to these variations and does not vary monotonically, as can be seen in the last column of the table. Somewhat surprisingly (c.f. Emanuel 1986, Rotunno and Emanuel 1987), the larger disturbances grow faster (also see discussion of variations of r_1, below). It would have been possible to vary r_1 to compensate for changes in L so as to leave the overall disturbance size constant. However, this would have required variations of $\eta^{(0)}$ (albeit very small ones) and it still would not have been possible to provide compensatory change of the updraft radius, r_0, which is much more important.

Our remaining comparisons require calculation of different linear solutions, rather than simply varying the parameters needed to evaluate the integrals as second order. As can be seen in table 7, increases in the overall disturbance size lead to decreases in the the threshold amplitude and increases in growth rate, though the effect is not strong. This was surprising in view of the results of Emanuel (1986)

and Rotunno and Emanuel (1987), which showed that the larger disturbances grew more slowly and very large ones barely intensifed at all. One significant difference between the size variation here and in Rotunno and Emanuel, is that here the size of the inner core of the disturbance was basically unchanged as the outer extent was varied, while in Rotunno and Emanuel the entire structure of the disturbance was scaled proportionally. Here, the primary reason for this increase in development potential for large disturbances is that the boundary layer formulation leads to greater entropy perturbations in the disturbance core for overall larger disturbances.

Table 8 examines variation of the upper level lateral eddy diffusivity with fixed lower level diffusivity. This also require varying the lateral diffusion ratio, K, which in turn changes the eigenvalue for the forcing parameter, $\eta_s^{(0)}$. Variations of the upper level lateral diffusion, $k_2^{(0)}$, add the further complication of changes to the length scale. The most significant change with increases in $k_2^{(0)}$ is the accompanying decrease in $\eta_s^{(0)}$, which is a first order change. The magnitude of $\eta_c^{(1)}$ also decreases and whether or not these compensate depends on the subcriticallity of the rest state. For example, if $\eta_a = 2.0$ then the critical velocity maximum is 0.91 m/sec ($= [\eta_a - \eta^{(0)}]fL/\eta_c^{(1)}$) for the K = 0.5 case and 0.80 m/sec for the K = 0.2 case, so the linear effect of decreasing $\eta^{(0)}$ dominates. However, if η_a is more subcritical at say 1.5, then the critical velocities are 1.2 m/sec and 1.4 m/sec, respectively, and the relative thresholds are reversed. Similar arguments apply to variations of $k_1^{(0)}$. However, since there is no additional help from variations in the length scale, unless the rest state is greatly subcritical, increases in lower layer diffusion make it more difficult to create an intensifying disturbance.

4. Superexponential Growth

The complex machinations above can be reduced to a simple amplitude equation. This will also serve to provide justification for the claim that superexponential growth is an appropriate description for the intensification rate of disturbances that are of sufficient amplitude to grow.

The quadratic homogeneous rate equation

$$\frac{da}{dt} = c_1 a + c_2 a^2 ,$$ (4.1)

where both c_j are constants, has known analytic solutions. There are two stationary points: one at $a = 0$ and another at $a = -c_1/c_2$. Figure 4 shows behavior with real coefficients $c_1 < 0$ and $c_2 > 0$ for several initial conditions. If $a_0 \equiv a(0) < 0$ then $a(t)$ decays from below. If $0 < a_0 < -c_1/c_2$ then $a(t)$ decays to zero from above. If $a_0 > -c_1/c_2$ then $a(t)$ blows up in finite time. This is faster than exponential, or *superexponential* growth.

Equating the critical amplitude from eq. (3.8) with the nonzero stationary point of eq. (4.1) implies

$$a_{crit} = \frac{\eta_a - \eta_s^{(0)}}{\eta_c^{(1)}} = -\frac{c_1}{c_2} > 0 .$$ (4.2)

Expansion of eq. (4.1) about a_{crit} with the perturbation amplitude $\tilde{a}$ defined by $a = a_{crit} + \tilde{a}$ leads to the linear amplitude equation $\frac{d\tilde{a}}{dt} = -c_1 \tilde{a}$. Comparing this with eq. (3.3) suggests equating $-c_1$ with $f\gamma^{(1)} a_{crit}$. With eq. (4.2), we then have the amplitude equation

$$\frac{da}{f\,dt} = \frac{\gamma^{(1)}}{\eta_c^{(1)}}(\eta^{(0)} - \eta_a)a + \gamma^{(1)}a^2 .$$ (4.3)

Though this has the behavior we desire for $a < 0$, the coefficients are not at all correct in that regime since the physics used in the determination of the coefficients

222

does not apply to a disturbance with a low level anticyclone near the origin. We may aspire to an analytic theory that exhibits the behavior of a cubic amplitude equation with a stable stationary point at greater amplitude than the unstable point that we have examined.

The specific solution, at least for positive a_0, is

$$a(t) = \frac{a_0}{a_0/a_{crit} + (1 - a_0/a_{crit})e^{a_{crit}\gamma^{(1)}t}}, \quad \text{for} \tag{4.4}$$

$$0 < t < t_{max} = \frac{1}{\gamma^{(1)}a_{crit}} \ln\left(\frac{a_0}{a_0 - a_{crit}}\right).$$

The concept of a simple growth rate cannot be applied to this type of solution. Clearly, if the initial amplitude equals the critical amplitude the singularity is never reached. However, if a_0 is only 20% greater than a_{crit}, with an organized surface velocity maximum of 2.5 m/sec, the singularity of infinite amplitude is reached in less than two weeks for the case examined in table 2 with $\eta_a \approx 1.5$. With a velocity maximum of 4 m/sec, the singularity is reached in a week. The mature amplitude would be reached more quickly. The paths of figure 4 roughly correspond to the parameter values we have examined if t is measured in seconds and the velocity amplitude is nondimensionalized with a velocity scale of about 8 m/sec. As can be seen, the time required to reach maturation decreases rapidly with increases in the supercriticality of the initial disturbance.

5. Discussion

5.1. Application to Given Disturbances

In section 3.3 it was shown that the finite amplitude stationary states found earlier in section 3.2 are unstable to perturbations with the same spatial structure. Such perturbations, rather than disrupting the structure, lead either to its intensification

or decay. States with amplitudes below a critical value decay, while states with larger amplitudes grow.

We defined earlier that the effective forcing parameter for a disturbance is $\eta_e \equiv \eta_s^{(0)} + \bar{a}\eta_c^{(1)}$. For cyclonic disturbances, $\bar{a}$ is positive and we found that $\eta_c^{(1)}$ is negative. The atmospheric ambient value of η_a, based only on the vertical thermodynamic structure, is determined mostly by the radiative-convective balance of the tropical atmosphere for the underlying sea surface temperature and the large scale atmospheric circulation. If the resulting η_e is less than η_a, then the disturbance will intensify. The appropriate eigenvalue for linear solutions is determined mostly by the internal friction ratios for the atmosphere, with slight dependence on the radiative decay time and the overall size of the test disturbance. For $\eta_s^{(0)} > \eta_a > \eta_e$, there is subcritical instability from a transcritical bifurcation with a linearly stable rest state.

Figure 5 shows a schematic of the decaying and growing states as a function of disturbance amplitude and the ambient value of the entrainment parameter, η_a. The theory presented here has determined the slope of the stability boundary (solid line) where it intersects the line of zero amplitude and the intercept, which is the linear instability criterion. We also know that there are no hurricane-like instabilities for $\eta < 1$, since that would lead to outflow rather than inflow in the lower troposphere and would not spin-up a low level cyclone. We might hope that the stability boundary resembles the dense line of dots shown in the figure. The sparse dotted line indicates what the stability boundary might look like if the curvature determined at third order were positive and very large; the calculation of this next order term would be an horrendous task. In that case the range of applicability of the theory presented here would be so small that an amplitude threshold would not be predicted. However, unless the nondimensional scaling was totally inappropriate or some important process totally omitted from the basic equations, the stability

boundary can probably be extended to near where a approaches 1 from below or η_s approaches 1 from above, whichever is reached first.

The unstable solutions we have found are not expected to be observed; they serve as dividers between decaying and intensifying disturbances. However, the actual disturbances one observes do not have the exact spatial form of the linear solutions found in Part I. In linear modes, the amplitudes of all measures of a disturbance evolve proportionately. When examining data from observed disturbances, which do not come in one eigenmode of only varying amplitude, different measures are not equivalent. Since the spatial structures of real disturbances are so variable from case to case, and so different from the solutions dividing decaying from growing states presented here, it is not clear when a given disturbance exceeds the threshold for growth.

To compare observed disturbances with this theory, the magnitude of the near surface vorticity maximum (with derivatives determined appropriate to the scale of the cloud cluster) may be the best choice as an amplitude. Strong surface winds with little vorticity on the cluster scale, do not force much convergence, and hence do not force much convection. Similarly, a strong surface pressure signal from a large diffuse system is not conducive to development. In the context of this theory, relative vorticity is a good indicator of the organized forced vertical velocity out of the boundary layer. This choice is also consistent with the observations of McBride and Zehr (1981) on intensifying disturbances.

5.2. Disturbance Height and Vertical Instability

Both Conditional Instability of the Second Kind (CISK) and Wind Induced Surface Heat Exchange (WISHE) rely on the presence of an organized *near surface* disturbance provided by some other mechanism. For Atlantic hurricanes the source

of these disturbances is primarily, though not exclusively, tropical easterly waves. These waves are thought to form as shear instabilities on the African easterly jet at about 70 kPa (Burpee 1972).

Conditions required for a disturbance to reach low levels have recently been examined by Miller (1990), who extended the results on vertical propagation of planetary waves found by Charney and Drazin (1961). Another possibility, raised by Miller, is that even evanescent disturbances may have strong near surface signals. He showed that trapped solutions may have a larger effect at low levels than propagating solutions, at least within one or two vertical e-folding scales of the disturbance source, i.e. the jet. A great encouragement for use of conditions that determine the vertical extent and propagation of disturbances in forecasting is that they can be evaluated from synoptic scale, rather than mesoscale, measurements.

When convergence is forced by large scale motions, the level of any convergence is crucial for the existence of CAPE with respect to the lifted parcels. The boundary layer is not well mixed with respect to moisture, and θ_e often drops several Kelvin in the lowest 5 kPa. A parcel rising 15 km with an additional 1 K of relative buoyancy for the entire depth of the troposphere provides about 500 J/kg of additional energy. Disturbances such as easterly waves that are strongest at about 70 kPa force convergence over some depth, however convergence above the boundary layer provides moisture and may lead to much precipitation but still provides little or no release of CAPE. In a weak cloud cluster, it has been found that the deepest convection, reaching to the tropopause, had the lowest cloud bases (at 450–600 m) and occurred in areas of large scale confluence (and assumedly convergence), though this was also in an environment where lesser clouds had already contributed to the moistening of the area (Warner et al. 1980). We expect that lower cloud bases should be associated with lower parcel origination levels and greater potential energy.

226

Any complete solution to the tropical cyclone intensification problem must confront the problem that there is a significant amount of CAPE that is allowed to build up without being released. This potential energy must be able to be released by a disturbance with a different energy source, e.g. barotropic or baroclinic. Possibly small changes in the vertical shear structure affect the vertical propagation of disturbances and only small amounts of barotropic or baroclinic energy are needed to generate the near surface cyclone that serves as a hurricane trigger.

5.3.   Objections to CISK and Replies

Emanuel (1986, 1989) has objected to CISK on several grounds. He noted (1986) that CISK is a linear theory and if such linear instability actually existed "weak tropical cyclones should be ubiquitous and not confined to maritime environments." Linear theory can not capture the knowledge that tropical cyclones arise out of rather strong preëxisting disturbances. He also argued that the assumed CAPE does not exist and that the requirement for "moisture convergence" attributed to CISK is insufficient. Even assuming that some CAPE does exist, Emanuel noted that the boundary layer in its undisturbed state is incapable of powering sufficient pressure drop for a mature tropical cyclone. CISK, as stated by Charney and Eliassen, makes no allowance for increasing the entropy of the boundary layer. Emanuel (1989) has recently gone further to state that Ekman pumping has a *negative* effect on development, which will be explained below. An additional objection, noted by the original proposers of CISK, is that without internal friction CISK suffers from one of the defects that it sought to surmount, namely that moist convective theory leads to fastest growth for the smallest spatial scales for many vertical heating profiles. Each of these objections will now be examined separately.

There is nothing inherently linear in the basic ideas of CISK (see discussion in Ooyama 1982). This work has been, in fact, devoted to presenting a finite amplitude

version of CISK that exhibits instability only to finite amplitude perturbations. At the time of the original CISK papers, no weakly finite amplitude instability theory had yet been published in the fluid dynamics literature. It is therefore not surprising that such techniques were not used. The theory in this work shows that one can have a CISK type theory with states that are linearly stable, but yet unstable to finite perturbations. Further, analysis of this finite amplitude system implies superexponential growth, consistent with the often observed explosive intensification. This is also consistent with the finite amplitude numerical analysis and interpretation of a similar model by van Delden (1989), which also exhibits greater intensification for stronger disturbances with the same forcing.

Emanuel has properly attacked the simplistic, and incorrect notion, that "moisture convergence", as proposed by Charney and Eliassen is sufficient to guarantee an energy source for intensification. Condensation and precipitation can be forced in a stable environment with moisture convergence. If the environment is moist stable such a process releases no additional energy supply and will have energetic cost to drive the motion and warm the atmoshere. The work of Ooyama and that presented here do not suffer from this defect since the cumulus parameterization is quite different. (This renders the term "CISK parameterization" absolutely confusing.) Boundary layer mass convergence is crucial to CISK, but the boundary layer air must have positive buoyancy with respect to air aloft for any net energy release to accompany covergence.

The most recent addition to Emanuel's objections is based on a simplified numerical model (1989). He claims that

> Ekman pumping first induces upward motion and adiabatic cooling.... The vortex core thus cools and θ_e decreases in the subcloud layer. Only when the lower-to-middle troposphere becomes nearly saturated can anomalous surface fluxes counter the drying effect of convective downdrafts to the extent that subcloud-layer θ_e actually increases. This

increase is associated with an increase in temperature aloft and thus with an amplification of the cyclone.

The process would only occur as outlined if convection initiated from Ekman pumping was primarily shallow and the lifted parcels did not become positively buoyant. However, observational evidence shows deep, heavily precipitating, convection in the regions of positive Ekman pumping (Cho and Ogura 1974) and detailed three dimensional numerical models also exhibit deep precipitating convection in locations with positive low level vorticity (Tuleya 1991).

The nonexistence of CAPE has not been established. There is, in fact, evidence for the existence of significant amounts of CAPE. Emanuel is correct that the undisturbed boundary layer has insufficient entropy to cause a pressure drop as great as is observed in a mature tropical cyclone. It is inescapable that the boundary layer entropy must be increased for a tropical storm to reach hurricane intensity. It does not necessarily follow that the boundary layer entropy must be significantly increased for the intensification process to begin.

Palmén (1948) showed that for September climatology in the North Atlantic tropics, a surface parcel lifted pseudoadiabatically had a large amount of relative buoyancy of order 6 K in the hurricane region. A similar parcel in February had little CAPE and a relative buoyancy of order 1 K. Kasahara (1954) and Bansal and Datta (1972) obtained similar results. So for many years, the existence of large amounts of summertime CAPE in the tropics was established wisdom.

This simple notion was challenged by Betts (1982). He showed that for some soundings from GATE and a composite hurricane sounding for a mature storm at about 50 km from the center, that the soundings were very close to a reversible moist adiabat (condensed liquid water was retained in calculating parcel buoyancies). However, tropical systems that do intensify rarely do so while passing through the GATE region and one would expect the sounding of a mature hurricane to indicate

229

the exhaustion of CAPE. Xu (1987; also see Xu and Emanuel 1989) examined this question more carefully for three tropical small island stations during the summer cyclone season: Truk in an area of moderate tropical cyclone activity; Koror in an area of frequent tropical cyclone activity; and Majuro with few storms in the vacinity (see climatology of Gray 1979). For Truk he found a shallow stable capping layer, with buoyancies of about 1 K for parcels lifted reversibly and about 5 K for pseudoadiabats in the middle troposphere. The errors were estimated at about 1 K, so the reversible path was considered to be neutral.[2] When only soundings unstable for reversible ascent were included, the buoyancies were found to be about 2 K, or double the standard deviation. Buoyancies were found to be largest at Koror and smallest at Majoro, which is consistent with the tropical cyclone climatology. For boundary layer parcels at Koror, the buoyancy for parcels lifted reversibly exceeded 2 K from 90 kPa on up and even reached 3 K. As noted earlier, even a 1 K surfeit over the depth of the troposphere provides more than 500 J/kg. Further, Williams and Renno (1991) have shown that if water loading is maintained for ascending parcels then consistency requires that the latent heat of fusion be included, at least at levels with temperatures below −20°C. This heat often more than compensates the energy lost to water loading, providing large amounts of CAPE for rising parcels, though Williams and Renno ignore the fact that when any of this ice precipitates, melting extracts this heat from the atmosphere, albeit at lower levels. It seems that even the data of Xu and Emanuel do not fully support their claim of neutrality.

Frank and Cohen (1989) performed numerical experiments with and without forced low level vertical velocities using their own cumulus parameterization. In the absence of forcing the resulting soundings closely followed the suggestion of Betts

---

[2]The analysis has the peculiar property, that the greater the variance of the data and the greater the assumed instrumental error, the stronger the support for the thesis of nearly neutral soundings.

by stabililizing nearly on a reversible moist adiabat. With forcing the final profiles were considerably more stable. That is, *convection initiated by large scale weak low level forcing is able to extract far more potential energy from the base state*. Hence, the reversible moist adiabat with full water loading may not be a relevant reference state in the presence of low level convergence.

It is only sensible to refer to a vertical profile as stable if it is stable to *undilute* ascent in the presence of whatever dissipative (but not entraining) processes exist. Even this is complex. A profile that is neutral with respect to nonprecipitating convection is unstable to precipitating convection if the water content of a parcel becomes less than what would result from adiabatic cooling (e.g. by loss of liquid water due to precipitation). However, the presence of liquid water in excess of what can condense locally (e.g. by rainfall into parcels, such as those near cloud base) can create a stable layer choking off convection. Even if the tropics were neutral for boundary layer parcels rising while retaining their water content, they are quite unstable for the precipitating convection that actually occurs during most of the summer and early autumn.

Objections to CISK based on length scale selection, or the lack thereof, are more subtle. In the context of the viscous theory presented here, there is some disturbance size that has the smallest amplitude requirement for growth, and this size is determined by the various frictional parameters. For the values of these parameters we have examined, the radius of maximum winds for the disturbance requiring the smallest amplitude for excitation, is a couple of hundred kilometers. This correspondence in scale with the actual initial disuturbances is fortuitous. The length scale dependence on frictional parameters is also weak, going only as the square root. For subcritical instabilities, the initial length scale is determined by the size of the initial disturbance. This is consistent with the observed large range of tropical cyclone sizes.

The WISHE theory of Emanuel is challenged by the recent simulations of Tuleya (1991). In a numerical experiment from initial conditions that led to a hurricane, when evaporation was fixed, preventing feedback between surface winds and boundary layer entropy, tropical storm strength winds were still obtained and during the early hours of the experiment the growth rate was only slightly lessened from the case with feedback. In the hurricane developing control run, during early development, the maxima in precipitation, low level vorticity production from vortex stretching, and upper level warming, were all in phase with the low level vorticity maximum, consistent with a CISK model of intensification.

6. Conclusions

This is the first analytic theory to examine the need for a finite amplitude disturbance for tropical cyclone intensification. The need for sizeable disturbances has long been recognized from observations and some numerical models. Nonlinear mathematical techniques have been used because linear theory is incapable of describing a phenomenon with an amplitude threshold for development.

The analysis was based on the $2\frac{1}{2}$ layer hurricane model developed by Ooyama (1969), which includes a simple diagnostic cumulus parameterization. Convection is only initiated by boundary layer convergence and is assumed to release available potential energy. The primary effect of the convection is to drive middle level inflow, i.e. entrainment. The inflow is, in turn, responsible for convergence of angular momentum. The primacy of this task is reflected in the form of the forcing parameter, which depends on the ratio of the potential boundary layer buoyancy to the middle troposphere stability. This is a measure of the ease with which convection drives inflow, rather than an absolute measure of CAPE. Warming is an indirect consequence through geostrophic adjustment, with the parameterization chosen, though probably

less so in nature.

Linear stationary solutions were found for a system with CISK type forcing and dissipative processes, including internal friction and radiative cooling. The heart of CISK is that the location of deep convection is determined by the large scale flow, through frictional convergence, and that this convection serves to intensify the large scale flow. The system was expanded to second order to determine the amplitudes of the st ionary solutions for a given amount of vertical instability. It was found that a stationary balance was maintained by decreasing the forcing for increases of amplitude. Hence, finite amplitude stationary solutions can exist in forcing regimes that are linearly stable. Further, these stationary states were found to be unstable. The amplitude of the stationary states therefore serves as a divider between growing and decaying disturbances for a given vertical stratification. The instability found here is a finite amplitude version of CISK.

There are several processes that contribute to the nonlinear balance at subcritical values of the forcing parameter. The following ones assist tropical cyclone intensification:

1. Since the surface stress increases faster than linearly with tangential velocity, the Ekman pumping also increases faster than linearly. Therefore, there is an accompanying more rapid than linear increase in the cumulus heating.

2. The presence of a finite amplitude disturbance, cyclonic at low levels and anticyclonic aloft, leads to advection by the disturbance of disturbance angular momentum in addition to the planetary angular momentum. In the lower level cyclone this serves to help increase the cyclonic flow. In the upper level anticylone this serves to weaken the resulting anticyclone, which decreases the energetic drain of the upper level flow on the overall system, and assists in cyclone development.

3. And in a vein similar to WISHE, the tangential velocities increase the surface entropy flux and can lead to an increase in the near surface entropy. The increased boundary layer entropy leads to more intense convection in the locations with positive Ekman pumping, which are also in phase with the positive temperature perturbations aloft.

It is clear that many of the objections to CISK have already been, or can be, defeated. The objection that CISK is a linear theory, and hence cannot show the need for a finite amplitude disturbance, has been quashed here (if one was not satisified by the numerical work in Ooyama 1969). CISK theories do require a reservoir of CAPE, but there is no need that the initial CAPE be sufficient to generate a mature storm, only that it be sufficient to overcome dissipative processes. In this work, as well as Ooyama's (1969) numerical integration, even an increase in the boundary layer entropy was incorporated into a CISK type theory. The naïve reliance on moisture convergence is not part of the formulation used here, nor was it used in any of Ooyama's work.

The finite amplitude nature of the instability here is tied to the existence of dissipative processes. In Part I we showed that even minimal internal dissipation prevents linear instability producing hurricane-like disturbances. However, even under conditions that are linearly stable, sufficiently intense disturbance initiated with a different mechanism are capable of extremely rapid intensification through CISK processes.

Appendices

A. The Linear System

The linearized, transformed, self-adjoint system for $v = (v_1, v_2, v_3)$ is:

$$r\frac{d}{dr}\left(k_2^{(0)}B\frac{drv^{(1)}}{r\,dr}\right) + k_s Arv^{(1)} = 0 \qquad (A.1)$$

where

$$B = \begin{pmatrix} KM - M - BK^2 + 1 - \eta^{(0)} & M - BK & 1 \\ M - BK & -B & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad (A.2)$$

$$A = \begin{pmatrix} \frac{M^2}{K} - M + BK - B & -\frac{M^2}{K} + B & -\frac{M}{K} \\ -\frac{M^2}{K} + B & \frac{M^2}{K} & \frac{M}{K} \\ -\frac{M}{K} & \frac{M}{K} & \frac{1}{K} \end{pmatrix}, \qquad (A.3)$$

with the nondimensional parameters $M \equiv \mu/k_s$, $K \equiv k_1^{(0)}/k_2^{(0)}$, and $B \equiv bf^2k_2^{(0)}/k_s^2$. This vector equation governs first order stationary and slowly varying solutions.

The transformation matrix used to reach this form, also used in the nonlinear calculations below, is

$$T = \begin{pmatrix} M - \frac{M}{K} - BK & M - BK & 1 \\ \frac{M}{K} - B & -B & 0 \\ \frac{1}{K} & 0 & 0 \end{pmatrix}. \qquad (A.4)$$

This form is unique up to multiplication throughout by a constant.

B. Details of the Nonlinear Problem

B.1. Expansions

The expansion of η_r requires the expansion of the v_i. The first order expansion of eq. (2.11) for χ_0, for stationary ($\gamma = 0$) or slowly varying ($\gamma^{(0)} = 0$, $\gamma^{(1)} \neq 0$) solutions, results in an inhomogenous equation for $\chi_0^{(1)}$:

$$\frac{\partial}{\partial r}\left(k\chi r \frac{\partial \chi_0^{(1)}}{\partial r}\right) = -\frac{C_\varepsilon |v_1^{(1)}|}{h_0}(\chi_s^{(0)} - \chi_0^{(0)})r + \frac{w^{-(1)}}{h_0}(\chi_0^{(0)} - \chi_1^{(0)})r . \tag{B.1}$$

This can be integrated using solutions for $v_1^{(1)}$ and $w^{-(1)}$ from the linear stationary problem. With the boundary conditions $\left.\frac{\partial \chi_0^{(1)}}{\partial r}\right|_{r=0} = 0$ and $\chi_0^{(1)}(r_1) = 0$,

$$\chi_0^{(1)} = \int_r^{r_1} \frac{1}{k\chi r'} \int_0^{r'} \left[-\frac{C_\varepsilon |v_1^{(1)}|}{h_0}(\chi_s^{(0)} - \chi_0^{(0)}) + \frac{w^{-(1)}}{h_0}(\chi_0^{(0)} - \chi_1^{(0)})\right] r'' dr'' dr' . \tag{B.2}$$

At lowest order, $\chi_2^{(0)}$ is fixed. The first order expansion from eq. (2.12) is

$$\chi_2^{(1)} = \Xi \theta_m^{(1)} . \tag{B.3}$$

With the linearization of an equation proportional to eq. (2.10) and the boundary condition $\chi_2^{(1)}(r_1) = 0$,

$$\chi_2^{(1)} = \int_r^{r_1} \frac{\Xi f}{c_p(\pi_1 - \pi_2)}(v_2^{(1)} - v_1^{(1)}) dr' . \tag{B.4}$$

The radiative cooling at lowest order, $Q^{-(1)}$, is proportional to $\chi_2^{(1)}$ and is obtained directly from this expression. The value of χ_1 was assumed constant at all orders.

At second order there are differences between the stationary and the slowly varying time dependent expansions for the vertical velocity out of the boundary layer with

$$\bar{w}^{(2)} = \frac{\bar{a}^2 \partial}{r \partial r} r \left[\frac{k_s}{f}v_1^{(2)} + \frac{C_D}{f}|v_1^{(1)}|v_1^{(1)} - \frac{k_s}{f^2}v_1^{(1)}\frac{\partial r v_1^{(1)}}{r \partial r}\right] \tag{B.5}$$

236

applying to stationary solutions and

$$\bar{w}^{(2)} = \frac{2\tilde{a}\tilde{a}k_s}{f}\frac{\partial rv_1^{(2)}}{r\,\partial r} + \frac{\tilde{a}\tilde{a}\gamma^{(1)}h_0}{f}\frac{\partial rv_1^{(1)}}{r\,\partial r}$$
$$+ \frac{2\tilde{a}\tilde{a}\partial}{r\,\partial r}r\left[\frac{C_D}{f}|v_1^{(1)}|v_1^{(1)} - \frac{k_s}{f^2}v_1^{(1)}\frac{\partial rv_1^{(1)}}{r\,\partial r}\right] \tag{B.6}$$

applying to slowly varying solutions. The spatial forms of $v_1^{(2)}$ and $v_1^{(2)}$ are not the same, hence the use of the diacritical marks above the (2) to distinguish them. Note that $\bar{v}_j^{(2)} = \tilde{a}^2 v_j^{(2)}$, but $\tilde{v}_j^{(2)} = 2\tilde{a}\tilde{a}v_j^{(2)}$. For later convenience when the linear and nonlinear terms will be included in separate integrals in creating the solvability conditions at second order, $w^{(2*)}$ is defined as $w^{(2)}$, but without the terms linear in the $v_j^{(2)}$, i.e.,

$$\bar{w}^{(2*)} = \tilde{a}^2\frac{\partial}{r\,\partial r}\left[\frac{C_D}{f}|v_1^{(1)}|v_1^{(1)}r - \frac{k_s}{f^2}rv_1^{(1)}\frac{\partial rv_1^{(1)}}{r\,\partial r}\right] = \tilde{a}^2 w^{(2*)} \tag{B.7}$$

$$\tilde{w}^{(2*)} = 2\tilde{a}\tilde{a}\frac{\partial}{r\,\partial r}\left[\frac{C_D}{f}|v_1^{(1)}|v_1^{(1)}r - \frac{k_s}{f^2}rv_1^{(1)}\frac{\partial rv_1^{(1)}}{r\,\partial r}\right] = 2\tilde{a}\tilde{a}w^{(2*)} . \tag{B.8}$$

Similar expansions can be performed for $\tilde{Q}^{+(2)}$, $\tilde{Q}^{+(2)}$, $\tilde{Q}^{+(2*)}$, $\tilde{Q}^{+(2*)}$, etc.

The diffusion terms will not be complicated further. For second order terms only v_j will be expanded with $k_j = k_j^{(0)}$ and $k_j^{(1)} = 0$, which ignores the dependence of k_j on h_j.

B.2. Weakly Nonlinear Equations

Integrations for function such as $Q^{-(1)}$, and all of the integrals of higher order terms needed later, were performed using a simple trapezoidal rule on a fixed mesh. The mesh consisted of 100 equal intervals from 0 to r_0 and 100 equal intervals from r_0 to r_1. The scheme is good to one part in 10^3 when comparing the integrals $\frac{1}{r}\int_0^{r'}\frac{drv_j}{r\,dr}r\,dr$ with $v_j(r')$ as a test. We define the operator $\Box^2 v \equiv \frac{d}{dr}\left(\frac{drv}{r\,dr}\right)$. Terms of the form $v_i\Box^2 v_j$ were first integrated once by parts.

Since the stationary equations can be easily derived from the time dependent system, only the latter is presented here. The coupled time dependent inhomgeneous equations at second order are

$$\tilde{a}\bar{a}\gamma^{(1)}h_1^{(0)}v_1^{(1)}r = 2\tilde{a}\bar{a}[k_1^{(0)}r\Box^2 v_1^{(2)} + \mu(v_2^{(2)} - v_1^{(2)})r + k_s v_3^{(2)}r] \qquad \text{(B.9)}$$
$$+ (\bar{a}\bar{a} + \tilde{a}\bar{a})\left[\zeta_1^{(1)}\frac{k_s}{f}v_3^{(1)}r + Q^{-(1)}(v_2^{(1)} - v_1^{(1)})r\right]$$

$$\tilde{a}\bar{a}\gamma^{(1)}\epsilon h_2^{(0)}v_2^{(1)}r = 2\tilde{a}\bar{a}[k_2^{(0)}r\Box^2 v_2^{(2)} + \mu(v_1^{(2)} - v_2^{(2)})r + k_s v_4^{(2)}r] \qquad \text{(B.10)}$$
$$+ (\tilde{a}\bar{a} + \bar{a}\bar{a})\left[\zeta_2^{(1)}\frac{k_s}{f}v_4^{(1)}r + Q^{+(1)}(v_1^{(1)} - v_2^{(1)})r\right]$$

$$\frac{\tilde{a}\bar{a}\gamma^{(1)}f}{g(1-\epsilon)}(v_1^{(1)} - \epsilon v_2^{(1)}) =$$
$$2\tilde{a}\bar{a}\left[\frac{k_s}{f}\Box^2 v_3^{(2)} - (\eta^{(0)} - 1)\frac{k_s}{f}\Box^2 v_1^{(2)} + bf(v_2^{(2)} - v_1^{(2)})\right] \qquad \text{(B.11)}$$
$$- \sigma\tilde{a}\bar{a}\frac{d\eta_c^{(1)}w^{(1)}}{dr} - \tilde{a}\bar{a}(\eta^{(0)} - 1)\frac{\gamma^{(1)}h_0}{f}\Box^2 v_1^{(1)}$$
$$- (\tilde{a}\bar{a} + \bar{a}\bar{a})\frac{d}{dr}(Q^{(2a)} - w^{(2a)})$$

$$\int_r^{r_1}\left(\frac{\partial\tilde{h}_1^{(1)}}{\partial t} + \epsilon\frac{\partial\tilde{h}_2^{(1)}}{\partial t}\right)r'\,dr' =$$
$$2\tilde{a}\bar{a}\frac{k_s}{f}(v_1^{(2)} + v_3^{(2)} + v_4^{(2)})r \qquad \text{(B.12)}$$
$$\frac{h_0}{f}\bar{a}\gamma^{(1)}\tilde{a}v_1^{(1)}r + 2\tilde{a}\bar{a}\frac{C_\rho}{f}\left|v_1^{(1)}\right|v_1^{(1)}r - 2\tilde{a}\bar{a}\frac{k_s}{f^2}\frac{\partial rv_1^{(1)}}{r\,\partial r}v_1^{(1)}r\,.$$

The first line of each of these four equations after the equals sign is similar to the terms of the first order stationary or slowly varying equations. For the time dependent equations, $\sigma = 1$. The remaining terms, including those on the left hand side are inhomogeneous.

The second order stationary equations can be easily obtained from eqs. B.9–B.12 by setting $\gamma^{(1)}$ to zero, $\sigma = 2$, and all of the tildes are changed to overbars (e.g. $v_1^{(2)}$ to $v_1^{(2)}$).

238

Before proceeding we will examine the integral on the left hand side of eq. (B.12). From the equations for geopotential perturbtion in Part I, $g\left(\frac{\partial h_1}{\partial t} + \epsilon\frac{\partial h_2}{\partial t}\right) = \frac{\partial \phi_1}{\partial t}$. We also have $\frac{\partial \tilde{\phi}_1^{(1)}}{\partial t} = \bar{a}\gamma^{(1)}\tilde{\phi}_1^{(1)}$. From the expansion of the gradient wind relation, $\frac{\partial \phi_1^{(1)}}{\partial r} = f v_1^{(1)}$. Integrating once, multiplying by r and again integrating leads to

$$\Phi_1^{(1)}(r) \equiv \int_0^r \phi_1^{(1)}(r')\, r'\, dr' \qquad (B.13)$$

$$= \int_0^r \left[\int_0^{r'} f v_1^{(1)}(r'')\, dr''\right] r'\, dr' - \frac{r^2}{r_1^2}\int_0^{r_1} \left[\int_0^{r'} f v_1^{(1)}(r'')\, dr''\right] r'\, dr'\ .$$

The integration constants have been determined by requiring that $\Phi_1^{(1)}(0) = \Phi_1^{(1)}(r_1) = 0$. These are the only conditions consistent with eqs. (B.9)–(B.12).

We now rework eqs. (B.9)–(B.12) into the vector form:

$$r\frac{d}{dr}\left(k_2^{(0)}B\frac{dr\tilde{v}^{(2)}}{r\, dr}\right) + k_* A r \tilde{v}^{(2)} = r T \mathcal{N}^{(2)}\ ,$$

where

$$\mathcal{N}^{(2)} = \bar{a}\bar{a}\eta_c^{(1)} \begin{pmatrix} 0 \\ 0 \\ \frac{f k_2^{(0)}}{k_*}\frac{dw^{(1)}}{dr} \end{pmatrix} \qquad (B.14)$$

$$+ \bar{a}\bar{a}\gamma^{(1)} \begin{pmatrix} h_1^{(0)}v_1^{(1)} \\ \epsilon h_2^{(0)}v_2^{(1)} - \frac{f}{gr}\Phi_1^{(1)} + h_0 v_1^{(1)} \\ \frac{f^2 k_2^{(0)}}{g(1-\epsilon)k_*}(v_1^{(1)}-\epsilon v_2^{(1)}) + (\eta^{(0)}-1)\frac{h_0 k_2^{(0)}}{k_*}\Box^2 v_1^{(1)} \end{pmatrix}$$

$$+ \frac{2\bar{a}\bar{a}}{\sigma} \begin{pmatrix} Q^{-(1)}(v_1^{(1)}-v_2^{(1)}) - \frac{k_*}{f}\zeta_1^{(1)}v_3^{(1)} \\ Q^{+(1)}(v_2^{(1)}-v_1^{(1)}) + \frac{k_*}{f}\zeta_2^{(1)}(v_1^{(1)}+v_3^{(1)}) + C_D|v_1^{(1)}|v_1^{(1)} - \frac{k_*}{f}v_1^{(1)}\frac{dr v_1^{(1)}}{r\, dr} \\ \frac{f k_2^{(0)}}{k_*}\frac{d}{dr}(Q^{(2*)}-w^{(2*)}) \end{pmatrix}\ .$$

To obtain $\mathcal{M}^{(2)}$, needed for the stationary problem, from $\mathcal{N}^{(2)}$, set $\gamma^{(1)}$ to zero, $\sigma = 2$, and all of the tildes are changed to overbars.

239

Acknowledgments

References

Anthes, Richard A. 1972: Non-developing experiments with a three-level asymmetric hurricane model; NOAA Tech. Mem. ERL NHRL-97, U. S. Dept. of Commerce, National Hurricane Research Laboratory, Coral Gables, Florida, Jan. 1972, 18pp.

Bansal, R. K. and R. K. Datta 1972: Certain aspects for intensification of tropical storms over Indian Ocean area; *Indian J. Meteor. Geophys.* **23** (4), 503–506.

Bergé, Pierre, Yves Pomeau, Christian Vidal 1986: *Order Within Chaos;* Wiley, New York, 329pp. Trans. by Laurette Tuckerman from the French *L'ordre dans le chaos* (1984), Hermann, Paris.

Betts, Alan K. 1982: Saturation point analysis of moist convective overturning; *J. Atmos. Sci.* **39** (7), 1484–1505. Also appeared in an earlier shortened form as 1982: Convective overturning and the saturation point, in E. M. Agee and T. Asai, eds., *Cloud Dynamics*, Reidel, Dordrecht, 117–133.

Burpee, Robert W. 1972: The origin and structure of easterly waves in the lower troposphere of North Africa; *J. Atmos. Sci.* **29** (1), 77–90.

Challa, Malakondayya and Richard L. Pfeffer 1990: Formation of Atlantic hurricanes from cloud clusters and depressions; *J. Atmos. Sci.* **47** (7), 909–927.

Charney, J. G. and P. D. Drazin 1961: Propagation of planetary scale disturbances from the lower into the upper atmosphere; *J. Geophys. Res.* **66** (1), 83–109.

Charney, Jule G. and Arnt Eliassen 1964: On the growth of the hurricane depression; *J. Atmos. Sci.* **21** (1), 68–75.

Cho, Han-Ru and Yoshimitsu Ogura 1974: A relationship between cloud activity and the low-level convergence as observed in Reed-Recker's composite easterly waves; *J. Atmos. Sci.* **31** (8), 2058–2065.

van Delden, A. 1989: On the deepening and filling of balanced cyclones by diabatic heating; *Meteorol. Atmos. Phys.* **41** (3), 127–145.

DeMaria, Mark and John D. Pickle 1988: A simplified system of equations for simulation of tropical cyclones; *J. Atmos. Sci.* **45** (10), 1542–1554.

Emanuel, Kerry A. 1986: An air-sea interaction theory for tropical cyclones. Part I: Steady-state maintenance; *J. Atmos. Sci.* **43** (6), 585–604.

Emanuel, Kerry A. 1989: The finite-amplitude nature of tropical cyclongenesis; *J. Atmos. Sci.* **46** (22), 3431–3456.

Emanuel, Kerry A. 1991: The theory of hurricanes; *Annu. Rev. Fluid Mech.* **23**, 179–196.

Frank, William M. and Charles Cohen 1989: Lapse rate adjustment in the tropical atmosphere; *Eighteenth Conference on Hurricanes and Tropical Meteorology, May 16–19, 1989 (extended abstracts)*, American Meteorological Society, Boston, Massachusetts, 159–160.

Gray, William M. 1979: Hurricanes: Their formation, structure and likely role in tropical circulation; in D. B. Shaw, ed., *Meteorology over the Tropical Oceans*, Royal Meteorlogical Society, Bracknell, Berkshire, England, 155–218.

Handel, Mark David 1990: *Tropical Cyclone Intensification From Finite Amplitude Disturbances, or How Hurricanes Hardly Happen*, Sc.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 261pp.

Handel, Mark David 1991: Tropical cyclone intensification. Part I: Impediment to linear growth, ms, Phillips Laboratory/GPAS, Hanscom AFB, Massachusetts, December, to be submitted to *J. Atmos. Sci.*

Handel, Mark David 1992: Tropical cyclone intensification. Part III: Development of an eye, ms, Phillips Laboratory/GPAS, Hanscom AFB, Massachusetts, December, to be submitted to *J. Atmos. Sci.*

Ince, E. L. 1926: *Ordinary Differential Equations;* reprinted in 1956 by Dover, New York, 558pp.

Kasahara, A. 1954: Supplementary notes on the formation and the schematic structure of typhoons; *J. Meteor. Soc. Japan* (Ser. II) 32 (2), 31–52.

Kleinschmidt, E., jun. 1951: Grundlagen einer theorie der tropischen zyklonen; *Archiv Meteor. Geophys. Bioklim.* A 4, 53–72. Unpublished translation by British Meteorological Office as "Principles of the theory of tropical cyclones".

Malkus, W. V. R. and G. Veronis 1958: Finite amplitude cellular convection; *J. Fluid Mech.* 4, 225–260.

McBride, John L. and Raymond Zehr 1981: Observational analysis of tropical cyclone formation. Part II: Comparison of non-developing versus developing systems; *J. Atmos. Sci.* 38 (6), 1132–1151.

Miller, Ronald Lindsay 1990: *Topics of Shear Instability: a: Viscous Destabilization of Stratified Shear Flow, b: Organization of Rainfall by an Unstable Jet Aloft*; Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 153pp.

Molinari, John and David Vollaro 1989: External influences on hurricane intensity. Part I: Outflow layer eddy angular momentum fluxes; *J. Atmos. Sci.* **46** (8), 1093–1105.

Montgomery, Michael T. and Brian Farrel 1991: Tropical cyclone formation, ms, Harvard University, Cambridge, Massachusetts, August.

Ooyama, Katsuyuki 1964: A dynamical model for the study of tropical cyclone development; *Geofis. Int.* **4** (3), 187–198.

Ooyama, Katsuyuki 1969: Numerical simulation of the life cycle of tropical cyclones; *J. Atmos. Sci.* **26** (1), 3–40.

Ooyama, K. V. 1982: On basic problems in theory and modeling of the tropical cyclone, in L. Bengtsson and J. Lighthill, eds., *Intense Atmospheric Vortices*, Springer-Verlag, 21–34. Reprinted with only minor changes as: Conceptual evolution of the theory and modeling of the tropical cyclone; *J. Meteor. Soc. Japan* (Ser. II) **60** (1), 369–380.

Palmén, Erik 1948: On the formation and structure of tropical hurricanes; *Geophysica* **3**, 26–38.

Reilly, Daniel H. and Kerry A. Emanuel 1991: Evidence of upper tropospheric triggering of tropical cyclogenesis, *Nineteenth Conference on Hurricanes and Tropical Meteorology, Preprints*, American Meteorological Society, Boston, Massachusetts, 202–205.

Riehl, Herbert 1948: On the formation of west Atlantic hurricanes; *Studies of the Upper-Air Conditions in Low Latitudes*, Part I, Miscellaneous Reports No. 24 of the Department of Meteorology of the University of Chicago, University of Chicago Press, 1-67.

Riehl, Herbert 1954: *Tropical Meteorology;* McGraw-Hill, New York, 392pp.

Rotunno, Richard and Kerry A. Emanuel 1987: An air-sea interaction theory for tropical cyclones. Part II: Evolutionary study using a nonhyrdrostatic axisymmentric numerical model; *J. Atmos. Sci.* **44** (3), 542-561.

Schlüter, A., D. Lortz and F. Busse 1965: On the stability of finite amplitude convection; *J. Fluid Mech.* **28**, 223-239.

Shapiro, Lloyd J. 1977: Tropical storm formation from easterly waves: A criterion for development; *J. Atmos. Sci.* **34** (7), 1007-1021.

Tuleya, Robert E. 1988: A numerical study of the genesis of tropical storms observed during the FGGE year; *Mon. Wea. Rev.* **116** (5), 1188-1208.

Tuleya, Robert E. 1991: Sensitivity studies of tropical storm genesis using a numerical model, *Mon. Wea. Rev.* **119** (3), 721-733.

Warner, C., J. Simpson, G. Van Helvoirt, D. W. Martin, D. Suchman, G. L. Austin 1980: Deep convection on day 261 of GATE; *Mon. Wea. Rev.* **108** (2), 169-194.

Williams, Earle and Nilton Renno 1991: The role of ice in the conditional instability of the tropical atmosphere, ms, Massachusetts Institute of Technology, Cambridge, Massachusetts, September, 49pp.

Xu, Kuanman 1987: Vertical structure and the convective characteristics of the tropical atmosphere; M.S. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 121pp.

Xu, Kuan-man and Kerry A. Emanuel 1989: Is the tropical atmosphere condtionally unstable?; *Mon. Wea. Rev.* **117** (7), 1471–1479.

Yamasaki, Masanori 1977: A preliminary experiment of the tropical cyclone without parameterizing the effects of cumulus convection; *J. Meteor. Soc. Japan* (Ser. II) **55** (1), 11–31.

Parameter Values and Eigenvalues

h_0	$= 1 \cdot 10^3$ m	c_p	$= 1005$ J/(kg·K)
$h_1^{(0)}$	$= 5 \cdot 10^3$ m	g	$= 9.8$ m/sec^2
$h_2^{(0)}$	$= 5 \cdot 10^3$ m	π_1	$= (.8)^{2/7} \approx 0.94$
ϵ	$= 0.4$	π_2	$= (.3)^{2/7} \approx 0.71$
f	$= 5.0 \cdot 10^{-5}$ sec^{-1}	Ξ	$= 1.2$
M	$= \mu/k_s = 0.5$	$\eta_s^{(0)}$	$= 2.75$
K	$= k_1^{(0)}/k_2^{(0)} = 0.2$	C_D	$= 2.0 \cdot 10^{-3}$
B	$= bf^2 k_2^{(0)}/k_s^2 = 5.0 \cdot 10^{-2}$	C_E	$= 2.0 \cdot 10^{-3}$
λ_1	$= 5.0 \cdot 10^3$ m^2/sec	$k_1^{(0)}$	$= \lambda_1 h_1 = 2.5 \cdot 10^7$ m^3/sec
λ_2	$= 6.2 \cdot 10^4$ m^2/sec	$k_2^{(0)}$	$= \epsilon \lambda_2 h_2 = 1.2 \cdot 10^8$ m^3/sec
μ	$= 2.5 \cdot 10^{-3}$ m/sec	b	$= 4.0 \cdot 10^{-6}$ sec/m
k_s	$= 5.0 \cdot 10^{-3}$ m/sec	k_χ	$= 4.0 \cdot 10^6$ m^2/sec
χ_s	$= 365$ K	r_0	$= 2.4 \cdot 10^5$ m
$\chi_0^{(0)}$	$= 352$ K	r_1	$= 9.5 \cdot 10^5$ m
$\chi_1^{(0)}$	$= 340$ K	L	$= (k_2^{(0)}/k_s)^{1/2} = 1.6 \cdot 10^5$ m
$\chi_2^{(0)}$	$= 344.$ K	V	$= fL = 7.9$ m/sec

Table 1: Values of parameters used, eigenvalues of the linear solution in figure 1, and resulting supplementary parameters. Several parameters are not independent.

Evaluation of Integrals

Ω	$(\text{m}^5/\text{sec}^3)$
$-\int_0^{r_0} k_2^{(0)} \left(\frac{dr v_1^{(1)}}{r\,dr}\right)^2 r\,dr$	$-1.8 \cdot 10^{10}$

Σ_i	$(\text{m}^5/\text{sec}^3)$
1. $\int_0^{r_1} \left[v_1^{(1)}\left(\text{M} - \frac{\text{M}}{\text{K}} - \text{BK}\right) + v_2^{(1)}\left(\frac{\text{M}}{\text{K}} - \text{B}\right) + \frac{v_1^{(1)}}{\text{K}}\right] Q^{-(1)}(v_1^{(1)} - v_2^{(1)})r\,dr$	$1.2 \cdot 10^{10}$
2. $-\int_0^{r_1} \left[v_1^{(1)}\left(\text{M} - \frac{\text{M}}{\text{K}} - \text{BK}\right) + v_2^{(1)}\left(\frac{\text{M}}{\text{K}} - \text{B}\right) + \frac{v_1^{(1)}}{\text{K}}\right] \frac{k_s}{f}\zeta_1^{(1)}v_3^{(1)}r\,dr$	$-3.6 \cdot 10^{10}$
3. $\int_0^{r_0} \left[v_1^{(1)}(\text{M} - \text{BK}) - v_2^{(1)}\text{B}\right] Q^{+(1)}(v_2^{(1)} - v_1^{(1)})r\,dr$	$-2.0 \cdot 10^{10}$
4. $\int_0^{r_1} \left[v_1^{(1)}(\text{M} - \text{BK}) - v_2^{(1)}\text{B}\right] \frac{k_s}{f}\zeta_2^{(1)}(v_1^{(1)} + v_3^{(1)})r\,dr$	$-1.7 \cdot 10^{10}$
5. $\int_0^{r_1} \left[v_1^{(1)}(\text{M} - \text{BK}) - v_2^{(1)}\text{B}\right] \left[C_D|v_1^{(1)}|v_1^{(1)} - \frac{k_s}{f}v_1^{(1)}\frac{dr v_1^{(1)}}{r\,dr}\right] r\,dr$	$1.6 \cdot 10^{10}$
6. $-\int_0^{r_0} k_2^{(0)}\eta_r^{(1)} \left(\frac{dr v_1^{(1)}}{r\,dr}\right)^2 r\,dr$	$-1.8 \cdot 10^{10}$
7. $-\int_0^{r_0} \frac{dr v_1^{(1)}}{r\,dr}\frac{fk_2^{(0)}}{k_s}\eta^{(0)}w^{(2*)}r\,dr$	$-4.6 \cdot 10^{10}$
8. $\int_0^{r_1} \frac{dr v_1^{(1)}}{r\,dr}\frac{fk_2^{(0)}}{k_s}w^{(2*)}r\,dr$	$1.9 \cdot 10^{10}$
9. $\int_0^{r_1} \frac{dr v_1^{(1)}}{r\,dr}\frac{fk_2^{(0)}}{k_s}Q^{-(2*)}r\,dr$	$-2.4 \cdot 10^{8}$

Γ_i	$(\text{m}^5/\text{sec}^2)$
1. $\int_0^{r_1} \left[v_1^{(1)}\left(\text{M} - \frac{\text{M}}{\text{K}} - \text{BK}\right) + v_2^{(1)}\left(\frac{\text{M}}{\text{K}} - \text{B}\right) + \frac{v_1^{(1)}}{\text{K}}\right] h_1^{(0)}v_1^{(1)}r\,dr$	$2.8 \cdot 10^{16}$
2. $\int_0^{r_1} \left[v_1^{(1)}(\text{M} - \text{BK}) - v_2^{(1)}\text{B}\right] \epsilon h_2^{(0)}v_2^{(1)}r\,dr$	$-5.6 \cdot 10^{15}$
3. $-\int_0^{r_1} \left[v_1^{(1)}(\text{M} - \text{BK}) - v_2^{(1)}\text{B}\right] \frac{f}{gr}\Phi_1^{(1)}r\,dr$	$1.6 \cdot 10^{13}$
4. $\int_0^{r_1} \left[v_1^{(1)}(\text{M} - \text{BK}) - v_2^{(1)}\text{B}\right] h_0 v_1^{(1)}r\,dr$	$8.3 \cdot 10^{14}$
5. $\int_0^{r_1} v_1^{(1)}\frac{f^2 k_2^{(0)}}{g(1-\epsilon)k_s}(v_1^{(1)} - \epsilon v_2^{(1)})r\,dr$	$5.1 \cdot 10^{13}$
6. $\int_0^{r_1} \frac{h_0 k_2^{(0)}}{k_s} \left(\frac{dr v_1^{(1)}}{r\,dr}\right)^2 r\,dr - \int_0^{r_0} \eta^{(0)}\frac{h_0 k_2^{(0)}}{k_s} \left(\frac{dr v_1^{(1)}}{r\,dr}\right)^2 r\,dr$	$-3.1 \cdot 10^{15}$

$\Omega = -1.8 \cdot 10^{10}\,\text{m}^5/\text{sec}^3$
$\Sigma = -9.0 \cdot 10^{10}\,\text{m}^5/\text{sec}^3$
$\Gamma = 2.0 \cdot 10^{16}\,\text{m}^5/\text{sec}^2$

$$\eta_c^{(1)} = -\frac{\Sigma}{\Omega} = -4.9$$
$$f\gamma^{(1)} = -\frac{\Sigma}{\Gamma} = 4.5 \cdot 10^{-6}\,\text{sec}^{-1}$$

Table 2: Values of integrals needed for both stationary and time dependent cases.

**Finite Amplitude Instability Criteria
With Varying C_D**

C_D	$\eta_c^{(1)}$	$f\gamma^{(1)}$ $(\sec^{-1})$
0.001	−4.2	$3.8 \cdot 10^{-6}$
0.002	−4.9	$4.5 \cdot 10^{-6}$
0.003	−5.6	$5.1 \cdot 10^{-6}$

Table 3: Variations in finite amplitude instability criteria with variations in bulk momentum diffusion coefficient, C_D. All other parameters are as in table 2.

**Finite Amplitude Instability Criteria
With Varying Coriolis Parameter**

f $(\sec^{-1})$	$\eta_c^{(1)}$	$f\gamma^{(1)}$ $(\sec^{-1})$	$-\frac{\eta_c^{(1)}}{fL}$ $(\sec/m)$
$5.0 \cdot 10^{-6}$	−2.7	$2.5 \cdot 10^{-6}$	3.4
$1.0 \cdot 10^{-5}$	−3.0	$2.8 \cdot 10^{-6}$	1.9
$3.0 \cdot 10^{-5}$	−4.1	$3.8 \cdot 10^{-6}$	0.87
$5.0 \cdot 10^{-5}$	−4.9	$4.5 \cdot 10^{-6}$	0.62
$7.0 \cdot 10^{-5}$	−5.3	$4.8 \cdot 10^{-6}$	0.48
$1.0 \cdot 10^{-4}$	−5.3	$4.8 \cdot 10^{-6}$	0.34

Table 4: Variations in finite amplitude instability criteria with variations in Coriolis parameter. The dimensional cooling parameter, b, is allowed to vary, so that the nondimensional cooling parameter, B, can be held constant. All other parameters are as in table 2, with the exception of V.

**Finite Amplitude Instability Criteria
With Varying Lateral Diffusion**

$k_1^{(0)}$ (m³/sec)	$\eta_c^{(1)}$	$f\gamma^{(1)}$ (sec⁻¹)	$-\frac{\eta_c^{(1)}}{fL}$ (sec/m)
$1.0 \cdot 10^7$	-3.5	$3.2 \cdot 10^{-6}$	0.70
$2.0 \cdot 10^7$	-4.4	$4.0 \cdot 10^{-6}$	0.62
$2.5 \cdot 10^7$	-4.9	$4.5 \cdot 10^{-6}$	0.62
$3.0 \cdot 10^7$	-5.5	$5.0 \cdot 10^{-6}$	0.63
$4.0 \cdot 10^7$	-6.8	$6.2 \cdot 10^{-6}$	0.68

Table 5: Variations in finite amplitude instability criteria with proportional variations in the lateral diffusion coefficients, $k_1^{(0)}$ and $k_2^{(0)}$, keeping K constant.

**Finite Amplitude Instability Criteria
With Varying Vertical Diffusion**

k_s (m/sec)	$\eta_c^{(1)}$	$f\gamma^{(1)}$ (sec⁻¹)	$-\frac{\eta_c^{(1)}}{fL}$ (sec/m)
0.003	-8.4	$4.6 \cdot 10^{-6}$	0.82
0.004	-6.1	$4.4 \cdot 10^{-6}$	0.69
0.005	-4.9	$4.5 \cdot 10^{-6}$	0.62
0.006	-4.2	$4.6 \cdot 10^{-6}$	0.58

Table 6: Variations in finite amplitude instability criteria with proportional variations in the linear boundary layer diffusion coefficient k_s and the vertical diffusion coefficient μ. The vertical diffusion ratio, M, is kept constant, so μ is also increasing as one moves down the table entries.

Finite Amplitude Instability Criteria
With Varying r_1

r_1	$\eta_c^{(1)}$	$f\gamma^{(1)}$ ($\sec^{-1}$)
6.0	−4.9	$4.5 \cdot 10^{-6}$
8.0	−5.1	$5.0 \cdot 10^{-6}$
10.0	−5.3	$5.2 \cdot 10^{-6}$

Table 7: Variations in finite amplitude instability criteria with variations in overall disturbance size, r_1. There is very slight variation in $\eta_s^{(0)}$ and r_0 between entries (see Part I).

Finite Amplitude Instability Criteria
With Varying Upper Level Lateral Diffusion

$k_2^{(0)}$ ($m^3/\sec$)	K	$\eta^{(0)}$	$\eta_c^{(1)}$	$f\gamma^{(1)}$ ($\sec^{-1}$)	$-\frac{\eta_c^{(1)}}{fL}$ ($\sec/m$)
$5.00 \cdot 10^7$	0.50	3.36	−7.5	$1.1 \cdot 10^{-5}$	1.49
$1.25 \cdot 10^8$	0.20	2.64	−6.4	$6.6 \cdot 10^{-6}$	0.80

Table 8: Variations in finite amplitude instability criteria with variations in the upper level lateral diffusion coefficient, $k_2^{(0)}$. The lateral diffusion ratio, K, is varied so as to keep $k_1^{(0)}$ constant. The cooling parameter is set at B = 0.005.

Figure 1: A solution to the linear stationary problem with M = 0.5, K = 0.2, B = 0.05, r_1 = 6.0. The velocities $v_1^{(1)}$ and $v_2^{(1)}$ are the upper and lower layer tangential velocities, respectively, and $v_3^{(1)}$ is proportional to the lower layer inflow velocity. The $v_j^{(1)}$ are all equal to zero at the origin and can be distinguished according to the key in the upper right corner. The line for $\frac{d r v_1^{(1)}}{r\,dr}$ has a positive value at the origin and is proportional to the Ekman pumping.

Figure 2: First order variations of the boundary layer and upper layer entropy, $\chi_0^{(1)}$ and $\chi_2^{(1)}$, respectively, for the linear solution in figure 1 and the parameters in table 1.

Figure 3: First order variation of the forcing parameter due to entropy changes, i.e. $\eta_r^{(1)}$, based on the $\chi_j^{(0)}$ of figure 2.

Figure 4: Orbits for the amplitude equation $\frac{da}{dt} = -1.5 \cdot 10^{-7} a + 5 \cdot 10^{-6} a^2$.

Figure 5: Schematic of the regions of decaying and growing states as a function of disturbance amplitude, a, and ambient forcing parameter, η_a. Points in the area labeled *decaying* evolve to the rest state, or if above the upper branch decay to a mature state. Points in the area labeled *growing* evolve towards the upper branch. The intercept on the horizontal axis gives the critical value of the forcing parameter for linear instability. The solid line shows the dependence of the threshold amplitude as a function of the ambient forcing parameter for small amplitude disturbances, as calculated in table 2. The dense dotted line shows a conjectured extension of that stability boundary, where the upper and lower branches are assumed to eventually connect at the left. (The conjectured upper branch would be determined by a different dominant balance than found in the solutions examined here.) The sparse dotted line indicates a possible behavior for a stability boundary that would greatly limit the usefulness of this type of theory, but is not thought to occur. Note the break in the amplitude (vertical) scale.

M = 0.50
K = 0.20
B = 0.05

r = 1.542

FIGURE 1

FIGURE 2

251
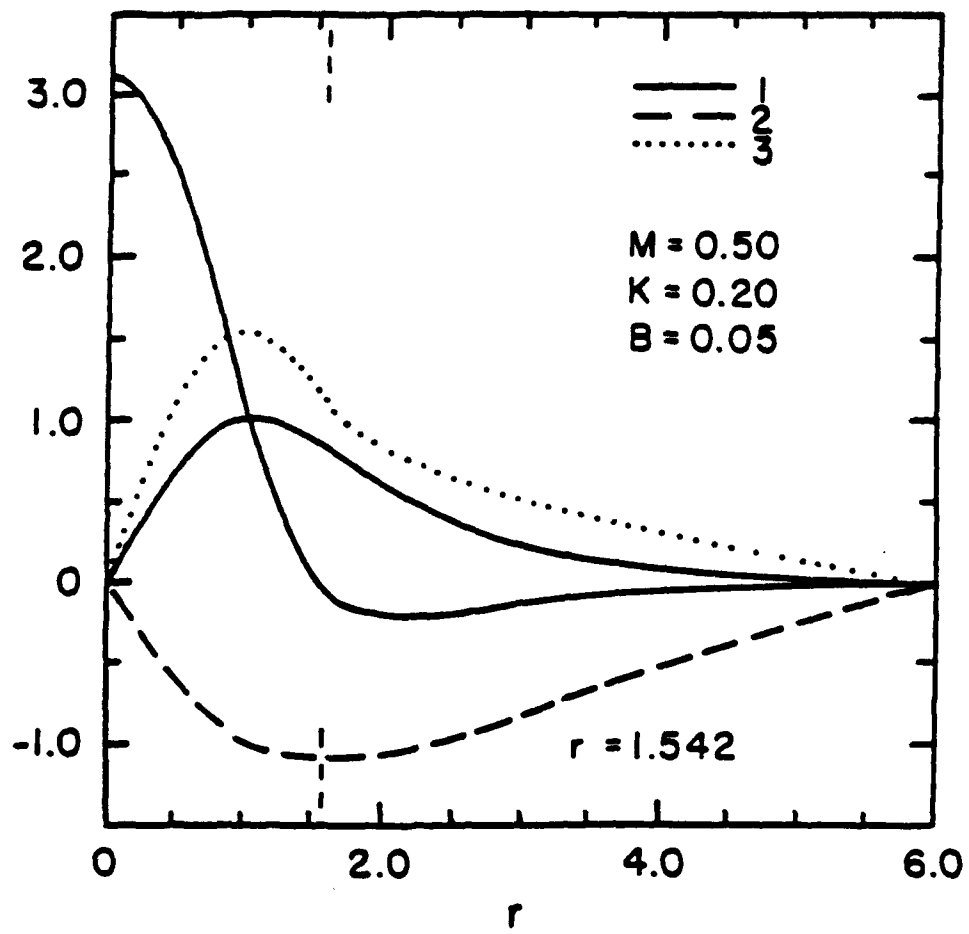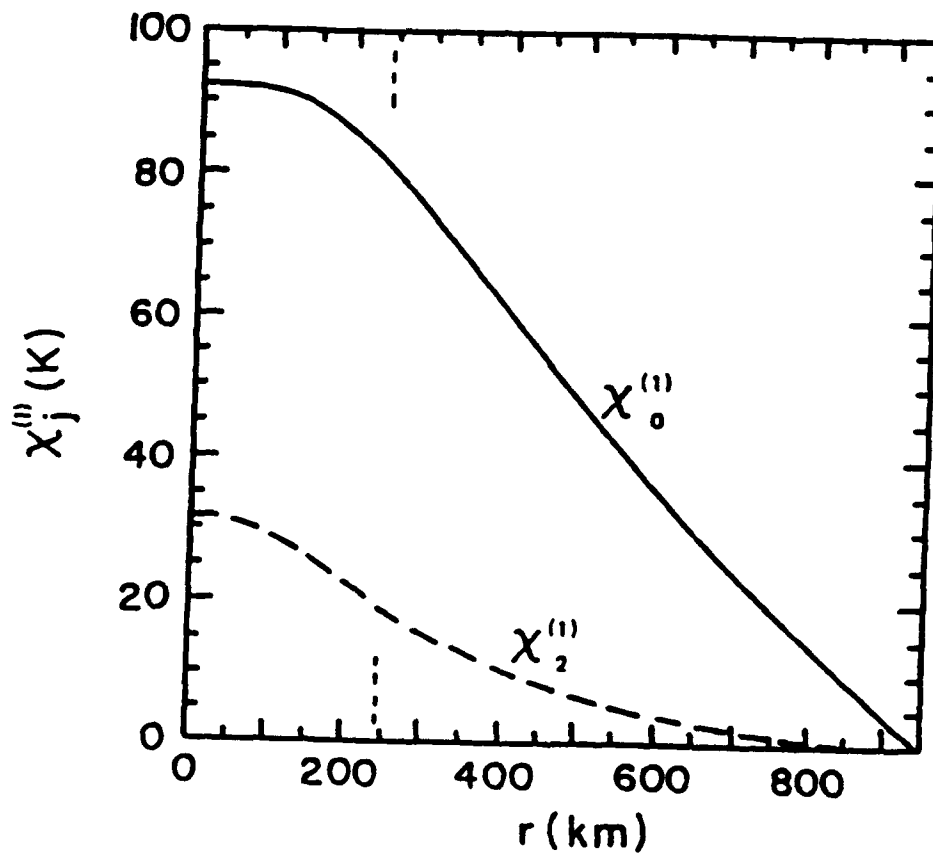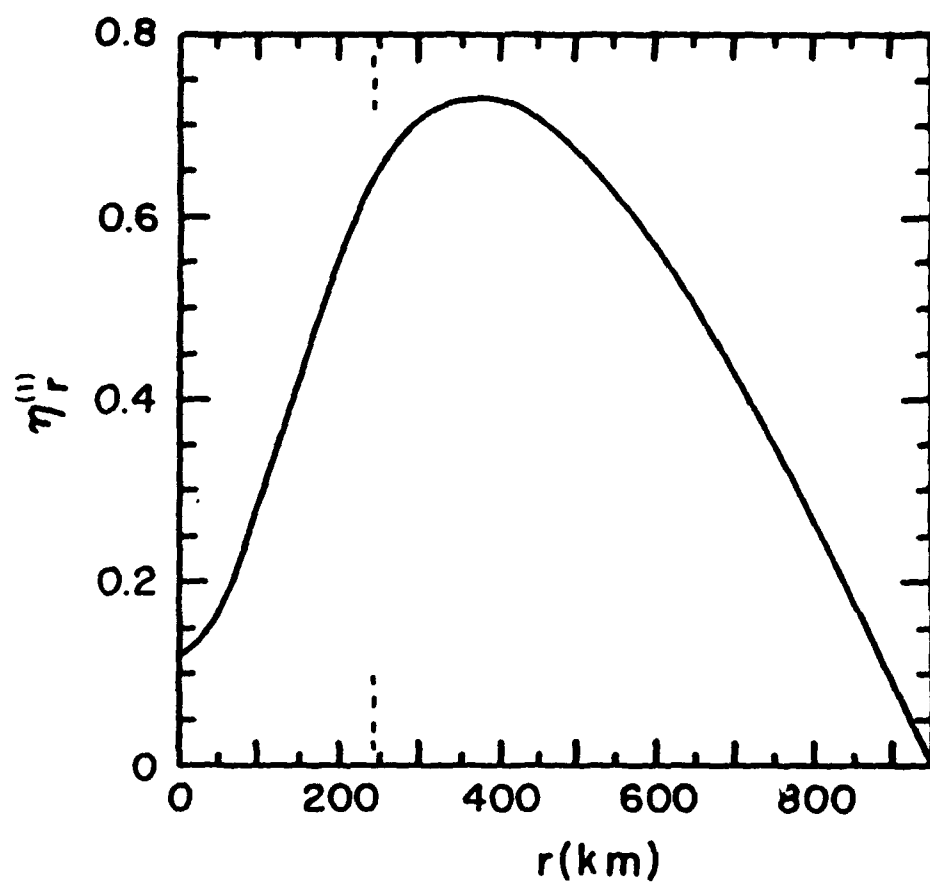
FIGURE 3

FIGURE 4

FIGURE 5

253

1991–1992 SCEEE Geophysics Scholar Program

Sponsored by

Department of the Air Force
Phillips Laboratory / Geophysics Directorate

Conducted by the

Southeastern Center for
Electrical Engineering Education

Final Report

Part IV. Climate Change Papers

Prepared by: Dr. Mark David Handel
Research Location: Phillips Laboratory / Geophysics Directorate
Phillips Laboratory Research Contact: Dr. J. William Snow
Date: May 29, 1992
Contract Number: F19628-86-C-0224

REFLECTIONS ON MORE THAN A CENTURY OF CLIMATE CHANGE RESEARCH

*Jamais, quels que puissent être le progrès des sciences, les savants de bonne foi et soucieux de leur réputation ne se hasarderont à prédire le temps.**

François Arago (1845)

Guest Editorial

When Joseph Fourier (1824) compared the action of the atmosphere on the Earth to that of glass covering a bowl, he was attempting to understand Earth's temperature in terms of the radiation balance between incoming solar radiation and outgoing dark radiation to space. When Svante Arrhenius (1896) calculated the effect of variations in atmospheric concentration of carbonic acid on the temperature of Earth's surface, he was seeking to explain how large climate changes could have occurred in the past. Arrhenius (1903) was also prescient to suggest that anthropogenic sources of carbon dioxide (CO_2) might lead to a warming in the future. However, as a Swede he did not consider this much of a problem, nor did he expect that the added CO_2 would be significant for several centuries. The process of assembling the following "Annotated Bibliography on the Greenhouse Effect and Climate Change" has made us aware of many trends in the study of climate. In the 1800s, climate studies were devoted to explaining the current climate and the great changes of the geological past. Recently, climate change research (or at least the on-paper justification) has focused almost exclusively on prediction of future climate change due to human activity and the repercussions of such climate change. We will examine some of the consequences of that focus shift and make some suggestions for future research agenda.

Water and Heat

The relative importance of water and carbon dioxide in controlling surface temperatures remains one of the most important unsettled scientific problems related to greenhouse change. This issue can be traced at least as far back as the 1860s, when John Tyndall argued for the overriding importance of water vapor while Gustav Magnus argued that there was significant infrared absorption by "dry air". Today the debate is over the sign of feedbacks involving water and how these interact with increasing CO_2. Accurate measurement of atmospheric water and its effects on the radiation balance of the atmosphere, should be made a top priority. We are encouraged by some movement in this direction.

With their excellent temporal and spatial coverage, satellites are being touted as the ultimate in measurement platforms for both water and temperature. However,

*Never, no matter what may be the progress of science, will honest scientists who care for their reputation venture to predict the weather.

255

such observations will not be trusted without *in situ* measurements using sensors significantly better than those presently on conventional radiosondes. This is particularly true for measurements of atmospheric water. More effort is needed on *in situ* measurements and on observational programs that can provide long time series of homogeneous data. Further, as new instruments are introduced they should be run in parallel with existing ones for longer periods of time than has been past practice. Too much of the existing data are considered suspect for climate studies because of calibration problems.

Despite the fact that most discussions of greenhouse change focus on the heat balance, changes in the hydrological cycle will have a much greater effect on humans and the biosphere than changes in temperature. The emphasis on temperature change is mostly due to a lack of confidence in our knowledge of water issues. Precipitation and atmospheric water have great variance on small spatial and temporal scales, most of which goes unmeasured. The lack of information on water processes can be traced to the daunting difficulty they present to anyone considering their study. Though changes of a few degrees in temperature might be an inconvenience, changes in water resources by a factor of two are a likely result of greenhouse change in many regions and could cause severe problems while exacerbating existing ones.

Science and Engineering

In the 1890s, Thomas Chamberlin examined climate change over geological time scales as part of an effort to understand Earth's formation and development. Chamberlin was already an exception to the increasing specialization in the sciences. There have been few reversals in this trend towards narrower fields of study. As perhaps the quintessential interdisciplinary research field, greenhouse climate change is reunifying the earth sciences, which had become extremely specialized. Study of the carbon cycle is forging links between the earth and life sciences. In addition, greenhouse climate change, with its deep links between human and natural systems, may provide one of the best opportunities for clarifying the methodologies connecting the natural and social sciences. We are impressed by the many scientists who have put in great effort to understand the work and methodologies of fields outside of their own. Many researchers now regularly confer with colleagues with whom they felt they had little in common only a few years ago.

Developing a background in specialties outside of one's own is difficult and time consuming. The following Bibliography came out of our efforts to educate ourselves. It was encouraged by many colleagues who either attempted to reduce our ignorance or wanted copies of our growing computerized reading list for themselves. It is a testament to the quality and vast quantity of literature relating to climate change and the greenhouse effect that much good work has been left untouched by the Bibliography. Still, we hope our compendium will make it easier

for at least a few workers to cope with the increasing interactions between the many disciplines that relate to climate change.

The rise in interdisciplinary activity around climate change research has led to the discovery of increasing complexity. The number and extent of interconnections are proving to be a rich source of surprises, making many simplifying assumptions difficult to justify. The increase in the accepted level of complexity underlies a real shift in research methodology to encompass a concern with the behavior of the complete climate system. This shift is one from a traditional scientific methodology to an engineering based methodology. The goal of science has typically been one of simplifying problems to understand processes; engineering deals with problems that one must solve practically in all their complexity.

The requirements of the policy and political communities, along with the public, for understanding the climatic consequences of increasing greenhouse gases have pushed many scientists into the unfamiliar roles of forecaster and engineer. The scientific community is at best unprepared and perhaps even unsuited for these tasks. Though short term weather forecasting is now largely an engineering problem, it has a sufficient experience base to use parameterizations that at least work, even if they can not fully be justified from first principles. We are far from being confident in the small scale parameterizations of climate models because of the lack of ability to confirm results. This problem is not likely to be solved in the near future.

The shift in research focus toward prediction of future climate has lead to the placing of increased emphasis on the modelling of climate by General Circulation Models (GCMs). A large fraction of all climate modelling efforts is now being devoted to simulations of climate with increased CO_2. Though we have learned much from past efforts of this type, these should be de-emphasized in favor of more detailed comparison of models with the present climate. The forefront of the modelling effort consists of models coupling the atmosphere and oceans. The present discordance between atmospheric and oceanic models when coupled, requiring large energy and water flux corrections at the interface, makes us suspicious of both types of models. Simulations with increased CO_2 provide little verification feedback to the modelling process. More detailed examinations should be made of changes associated with the seasonal cycle, diurnal cycle, volcanic aerosols, industrial aerosols, and land use changes. Further, energy fluxes should be examined more closely in models, since the fluxes better reflect the physical processes than state variables (such as temperature). Simpler models, such as two dimensional ones, should not be neglected. These are useful for understanding the climate system and identifying important processes, even if they can not produce the grail of accurate regional forecasts. In all their complexity, GCMs are often as difficult to fathom as the real atmosphere. For the engineering approach using complex models to progress, increased emphasis is needed on the scientific approach of understanding processes in a simpler context.

Mark David Handel and James S. Risbey

Uncertainty and Equity

Despite all the uncertainties, a consistent scientific picture seems to be emerging. The observed mean surface temperature increase over the past century has been smaller than most sensitivity estimates of the climate to increasing greenhouse gases alone would predict. However, inclusion of increases in Northern Hemisphere industrial aerosols in the radiation balance plausibly accounts for the discrepancy (Charlson 1992). Observations indicate that the increase in Northern Hemisphere temperature has been mostly at nighttime (Karl *et al.* 1991). This fits a picture of anthropogenic climate change from increasing both greenhouse gases and aerosols. The "consensus" estimates of climate sensitivity to CO_2 are consistent with the paleoclimate records of temperature and CO_2 concentration. If it were not for the policy ramifications, the scientific results on greenhouse change on the whole appear orderly and well founded, especially when compared with the levels of agreement and consistency of many other scientific problems.

Some pundits rely with great faith on a favorite economic model to forecast great economic cost and disruption for reducing greenhouse gas emissions while at the same time decrying the quality of climate models. Rind *et al.* (1988) have noted that the foundations of the climate models are far firmer than those of the economic models. For small perturbations to the climate system (say less than 5 K), not leading to dramatic qualitative changes to the atmospheric or oceanic circulations, existing atmospheric GCMs or coupled models are useful for examining climate trends on continental spatial scales and decadal temporal scales over the next century. Projections on smaller scales and analyses of possible radical changes, desired by many in the policy community, are not likely to be trustworthy for some time. Other than an awareness that hydrological processes are very sensitive to even small climate changes, there is little knowledge of how water resources are likely to change with global warming. Even if existing models are largely correct, we are unlikely to believe forecasts of dramatic changes very far from our experience. This is a worrisome attribute of human nature.

Issues of scientific uncertainty are most often raised when considering mandating changes in economic activity to protect the environment. The costs of mitigating climate change do not fall on the same individuals or at the same time as the costs of adapting to climate change. Some macroeconomic studies have contended that reducing greenhouse gas emissions will be very costly, perhaps even more costly than adapting to climate change. Other studies have contended that there will be net benefits from great reductions in our energy consumption even in the absence of climate change. Regardless, we should be wary of analyses that only examine effects on gross product but ignore equity issues. In the framework of macroeconomic theory, bank robbery is a zero sum activity. Microeconomic studies on the effects of climate change on specific sectors, locales, and groups, as well as studies on the economics of energy usage, are of more immediate value. Just as in climate research, we feel it is important in economics to emphasize process studies

and small scales. In particular, further analysis is needed of the barriers to energy technology substitutions and conservation, the costs of possible climate change, ways of arranging transfer payments to those who bear those costs but gain few of the benefits of activities leading to greenhouse gas emissions, and the changes in living conditions associated with alternative policy strategies. Studies that mostly emphasize changes in growth of gross product are as useful as climate forecasts that focus only on changes in the global mean temperature.

Most greenhouse gas emission scenarios, with their resultant climate change and economic projections, are not forecasts; they are decision making tools. There are fatalists who feel that our course is set, and some who feel they can predict it. However, we prefer to believe that the information we are accruing stands some chance of influencing future human behavior. Some millenia from now, as Arrhenius (1908) and Callendar (1938) suggested, we may even use greenhouse gases to prevent the next ice age.

Acknowledgments

References

Arago, François 1845: "La prédiction du temps", in M. J.-A. Barral (ed.) 1858: *Œuvres de François Arago*, vol. 8, Notices Scientifiques vol. 5, 1–24, Paris.

Arrhenius, Svante 1896: On the influence of carbonic acid in the air upon the temperature of the ground, *Philos. Mag.* (Ser. 5) 41 (251), 237–276.

Arrhenius, Svante August 1903: *Lehrbuch der kosmischen Physik*, Hirzel, Leipzig, 2 vols., 1026pp. (See vol. 2.)

Arrhenius, Svante 1908: *Worlds in the Making*, trans. by H. Borns, Harper, New York, 230pp. (See chap. 2.)

Callendar, G. S. 1938: The artificial production of carbon dioxide and its influence on climate, *Q. J. R. Meteor. Soc.* 64 (275), 223–240.

Chamberlin, T. C. 1897: A group of hypotheses bearing on climatic changes, *J. Geol.* 5 (7), 653–683.

Chamberlin, T. C. 1898: The influence of great epochs of limestone formation upon the constitution of the atmosphere, *J. Geol.* 6 (6), 609–621.

Chamberlin, T. C. 1899: An attempt to frame a working hypothesis of the cause of glacial periods on an atmospheric basis; parts I, II, III; *J. Geol.* 7 (6), 545–584; *J. Geol.* 7 (7), 667–685; *J. Geol.* 7 (8), 751–787.

Charlson, R. J., S. E. Schwartz, J. M. Hales, R. D. Cess, J. A. Coakley, Jr., J. E. Hansen, and D. J. Hofmann 1992: Climate forcing by anthropogenic aerosols, *Science* 255 (5043), 423–430.

Fourier, Jean-Baptiste Joseph 1824: Remarques générales sur les températures du globe terrestre et des espaces planétaires, *Ann. Chim. Phys.* (Ser. 2) 27, 136–167. Reprinted as Memoire sur les températures etc. in 1827: *Mem. Acad. R. Sci. Inst. Fr.* (Ser. 2) 7 [often listed under *Academie des Sciences* (Paris)], 569–604.

Karl, Thomas R., George Kukla, Vyacheslav N. Razuvayev, Michael J. Changery, Robert G. Quayle, Richard R. Heim, Jr., David R. Easterling, and Cong Bin Fu 1991: Global warming: Evidence for asymmetric diurnal temperature change, *Geophys. Res. Lett.* 18 (2), 2253. Correction in 1992: *Geophys. Res. Lett.* 19 (2), 219.

Magnus [Gustav] 1863: On the diathermancy of dry and moist air (trans. from German), *Philos. Mag.* (Ser. 4) **26** (172), 21–30.

Rind, David, Arthur Rosenzweig, and Cynthia Rosenzweig 1988b: Modelling the future: A joint venture, *Nature* **334** (6182), 483–486.

Tyndall, John 1863b: On the passage of radiant heat through dry and humid air, *Philos. Mag.* (Ser. 4) **26** (172), 44–54.

SCEEE Geophysics Scholar Program   MARK DAVID HANDEL
PL/GPAS
Hanscom A.F.B., MA 01731, U.S.A.

Center for Meteorology and Physical Oceanography   JAMES S. RISBEY
Massachusetts Institute of Technology, Room 54-1626
Cambridge, MA 02139, U.S.A.

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

LIGHTNING RADAR REFLECTIVITY: 0.1 m^2/m DURING CURRENT, 0 m^2/m BETWEEN STROKES

Prepared by: Dr. Stan Heckman

Research Location: Phillips Laboratory,
                                      Geophysics Directorate
                                      Hanscom AFB, MA 01731

AF Research Colleague: Arnold A. Barnes, Jr.

Contract Number: F19628-86-C-0224

SECTION 1

Motivation:

I would like to compute energetic constraints on lightning leader propagation. I would also like to compute stability conditions for that propagation. Both of these computations require an understanding of how rapidly the leader cools. Were this cooling due to conduction alone, we could compute it, given channel radius. But time-resolved photography (Vince Idone, private communication) strongly suggests that lightning channels are deformed by turbulence, and hence, cooled by turbulent convection. The only measurements of lightning channel cooling rate are measurements of the rate of decay of radar echoes from lightning.

Conclusions:

Radar echo decay rates vary from .1 dB/ms to 5 dB/ms. These decay rates are almost certainly due to decay of electron number density (other explanations don't work). The decay of electron number density is almost certainly due to decay of temperature (other explanations don't work). 0.1 dB/ms to 5 dB/ms implies 3 to 120 kelvins/ms. These decay rates require either surprisingly small turbulent diffusivities ($< .01$ m^2/s) or surprisingly large electric fields in the decaying lightning channel.

What's in the rest of this section:

A summary of past measurements and my interpretation of those measurements.

Lightning radar reflectivity: $0.1 \ m^2/m$ during current, $0 \ m^2/m$ between strokes

With a 12 GHz CW radar, we viewed 1.5 m of a triggered lightning channel. Between strokes, radar echoes fell to less than $1 \ cm^2$, indicating channel conductivity less than one siemens. During current, the 1.5 m segment had a cross section between $200 \ cm^2$ and $1700 \ cm^2$, which, naively interpreted, requires an overdense plasma 1 to 11 centimeters in diameter. In strokes followed by long continuing current, reflectivity remained relatively constant for tens or hundreds of milliseconds, then fell faster than 0.2 dB/ms. In strokes not followed by long continuing current, reflectivity began decaying immediately, at 3 to 5 dB/ms. These echo decay rates imply 3 to 30 K/ms temperature decay.

1. HISTORY

Browne [1951] made the first attempt to interpret echo brightness; he imagined a line of electrons perpendicular to the beam, each electron uninfluenced by the motion of its neighbors. He computed that he needed $5 \ 10^{15}$ electrons per meter to reproduce his radar signal.

Ligda [1956] reminded us that electrons ignore neighbors only if there are few neighbors (underdense plasma). With sufficient electron density (overdense plasma), the electrons collectively reflect almost all the incident energy. Because further increases of electron density cannot further increase the radar reflectivity, radar cannot determine the electron density, it can only provide a lower bound. Ligda imagined the lightning as at least one channel, with a radius of at least two centimeters.

Hewitt [1957], with radar equipment designed for the study of lightning, measured reflectivity versus distance each millisecond. He observed some echoes decaying a few milliseconds after a stroke, while other echoes endured for tens of milliseconds, then dropped 20 dB in the ten milliseconds immediately preceding a subsequent return stroke. Rapidly decaying echoes most often occurred at low elevation; enduring echoes, at high elevation. The radar reflecting region grew about 20 km/ms in the inte. val between strokes.

Hewitt imagined the rapidly decaying echoes to be cooling channels of the previous return stroke, and the enduring and extending echoes to be new channel growth inside the cloud. He reminded us that collisions reduce the radar reflectivity, and that collision frequency is an increasing function of electron energy. Hewitt assumes a field inside the lightning arc of 3 MV/m, which in 1957 implied a collision frequency of $4.5 \ 10^{11}$ Hz. Hewitt computed that this large collision frequency should cause 99% absorption of his radar signal. Assuming that a 300 meter long section of channel will contain 50 meters reflecting in phase, Hewitt computed that $2.4 \ 10^{14}$ electrons per meter would reproduce his radar signal. Assuming a channel diameter of 2 cm, he derived a volume density of $2.5 \ 10^{18}$ electrons per meter cubed. He knew that, without some ionizing process, electron ion recombination would reduce this density more than three orders of magnitude in a millisecond. His lightning echoes last tens of milliseconds. Hewitt imagines electron density to be maintained by a steady current in the new channels during the interstroke interval.

Atlas [1958] offers a different explanation for the long duration of lightning echoes: a large (cloud sized) volume of low electron density plasma as the radar target. (Malan proposed the same geometry in 1937). Since Atlas, like Hewitt, assumes electrons are removed only by recombination with positive ions, he had the recombination rate equal to the product of electron density and positive ion density. He took the positive ion density equal to the electron density (no negative ions) and computed that 10^{12} to 10^{16} electrons per cubic meter would require 100 milliseconds to 100 seconds to decay to one tenth their initial number, in agreement with his observations.

Dawson [1972] offers a third explanation for the long duration of lightning echoes; the electron density is in thermodynamic equilibrium, and the channel requires tens of milliseconds to cool. Dawson considers the magnitude of Browne's echo, and discovers that no long thin channel could produce an echo so strong, even if the lightning plasma were as reflective as a perfect conductor at right angles to the beam. Dawson concludes that radar is reflected from an initially overdense highly branched streamer system.

Dawson suggested that future studies might measure the initial expansion of the lightning channel by examining the initial rise to peak value of the radar signal, and might measure the channel radius by measuring echo duration and comparing to theoretical predictions of cooling rate as a function of radius.

Cerni [1976] looked for cloud to ground lightning echoes with a ten centimeter radar. Cerni's eyes detected many cloud to ground flashes in the radar beam, but his radar detected none of them. Cerni did see three intracloud echoes, much weaker than Dawson predicted for overdense echoes.

Holmes et al [1980] find echoes at 11 cm always orders of magnitude less than Dawson predicted for overdense channels, and never find any echoes at all at 3 cm. Both findings are expected if lightning channels are underdense.

Proctor [1981] examined lightning with 3 cm, 5.5 cm, 50 cm, and 111 cm radar. The 50 cm and 111 cm radars detected every flash visually observed in the beam. At 5.5 cm, the radar detected only a few of those flashes observed, and at 3 cm, the radar saw no lightning at all. At 5 cm, Proctor's echoes (usually) vanished in milliseconds, while at VHF, echoes persisted for tens of milliseconds. These results are expected if lightning is (usually) underdense at microwave frequencies, but always overdense at VHF. Proctor also made simultaneous measurements with identical beamwidths, and found a lightning radar cross section twice as great at 111 cm as at 50 cm. Finally, Proctor made simultaneous measurements with overlapping beam volumes, one twenty times the size of the other, and found only a factor of six more reflectivity in the larger volume.

Mazur et al [1985] argue against Cerni, Holmes, and Procter that lightning is overdense. Mazur et al's echoes at 10 cm look like Hewitt's and Proctor's at VHF; that is, the radar cross section is relatively constant for 25 to 200 milliseconds before beginning a 0.2 dB/ms decay. This is expected if lightning is initially overdense. So long as the channel remains overdense, the radar echo will remain independent of temperature and electron density. When the channel finally cools to an underdense plasma, the radar echo begins to decay. Mazur et al calculated that, to reproduce the 0.2 dB/ms decay observed, the hot channel diameter must exceed 8 cm.

Williams et al [1989] reason that lightning is an arc, arcs are hot, hot air is overdense at all radar wavelengths. They attribute Holmes' and Proctor's inability to see lightning at 3 cm to the great brightness of precipitation at 3 cm. Williams et al offer a model of the radar cross section of a long, thin, tortuous dendrite of overdense plasma, predicting radar cross sections much smaller than Dawson, and compatible with the small cross sections observed by Holmes.

Krehbiel et al [1991] observed radar echoes associated with K-changes. These echoes began decaying immediately, at 0.5 to 2 dB/ms. The radar reflectivity dropped into the noise (3 to 6 dB below peak echo) between K-changes. The same flash containing these K-changes includes one echo that requires almost 200 ms to fall the 3 to 6 dB into the noise.

2. RADAR PICTURE OF LIGHTNING

These radar observations suggest that lightning is a bushy ended dendrite, several kilometers in extent, with branches separated from each other by hundreds of meters. The observations also suggest that between strokes, some branches grow at 10-100 m/ms, and some channels cool and become non-conductive.

We imagine a dendrite. *Malan* [1937] and *Atlas* [1958] imagine instead a diffuse blob of cold plasma. Of course, Malan and Atlas have seen that lightning from cloud to ground is dendritic. Malan and Atlas suggest that at high altitudes lightning changes from dendritic arc to diffuse glow. Atlas imagined such a diffuse blob because he needed a low electron density to fit his theory for the long duration of his radar echoes, and because, on his PPI display, radar echoes did look like cloud sized blobs. Subsequent observations with more modern recording tools see lightning a single range cell in extent, much smaller than cloud size. And more recent theory for the long duration of lightning echoes does not require small electron densities. Finally, underdense plasmas scatter power proportional to the fourth power of wavelength, so an underdense blob should be 16 times brighter at 11 cm than at 5.5 cm, 24 times brighter at 111 cm than at 50 cm, and two million times brighter at 111 cm than at 3 cm. This is incompatible with the observed wavelength dependence of lightning. (See figure 9.) We conclude that lightning is commonly a thin dendrite of hot plasma, not a volume filling blob of cold plasma. If lightning is ever a diffuse glow, this has not been seen by radar.

We imagine bushy ends. Proctor, at 5 cm, saw single echoes that moved rapidly in range, erasing their reflectivity behind them. At VHF he instead saw echoes that persisted and grew in extent, becoming complex assemblies of echoes (dendrites). When Proctor increased the gain of his 5 cm radar, he saw echoes

that looked like his VHF echoes. We imagine that at 5 cm, without increased gain, Proctor could only see the bushy leader heads, the rest of the channel was hidden in the noise.

We imagine only a few limbs, separated by hundreds of meters. Dawson, Mazur et al, and Williams imagine more closely spaced limbs. *Dawson* [1972] proposed that *Browne's* [1951] echo was from a complex lightning dendrite, with many branches per radar resolution volume. *Mazur et al* [1985] allowed that their reflectivities *might* be due to many limbs of the lightning channel crossing the same radar resolution volume. *Williams et al* [1989] tentatively concluded that the lightning dendrite was a volume filling target, with several limbs per radar resolution volume, because the mean extent of their echoes (6.7 km) exceeds their radar beamwidth (0.5 - 2.5 km). We disagree, imagining only zero or one channel per radar resolution volume. Were there many limbs per radar resolution volume, then the extent of the lightning echo would be the size of the 3D envelope containing the volume filling thicket of lightning channels, almost independent of radar beamwidth. If, on the other hand, there were only a single, thin, more or less straight limb crossing the volume at some random angle, then the extent of the limb within the radar beam would be linearly proportional to the width of the beam. Holmes et al [1980], with a beam diameter of several hundred meters, saw less than 300 m extent for half of all flashes, and less than 2 km for the other half. *Mazur et al* [1985], with similar beam diameter, saw only 3 of 40 flashes extend into two adjacent 300 m range bins. (4 of 40 channels intersected the beam at two separated positions, 1.2, 1.2, 0.9, and 3.3 km apart). *Zrnic et al* [1982], with the same beam as Mazur et al, saw three quarters of the flashes extending into adjacent beam volumes, with extents ranging from 600 m to 3 km. *Hewitt* [1957], with a beam diameter of two or three kilometers, saw typical horizontal extents of 1500 to 2000 km. (He inferred vertical extents of 3 to 6 km by a unique method.) *Williams et al* [1989], with a beam diameter of 500 m to 2500 m, saw extents from 0 to 35 km, 4 km being typical. *Ligda* [1956], with a scanning beam, effectively many kilometers beam diameter, saw one echo with extent greater than 100 miles. In general, larger beam diameters saw larger lightning extents, so we conclude that the lightning dendrite is made of only a few limbs, typically only one through any radar resolution volume.

We imagine the lightning channels extending into virgin air between strokes at 10 to 100 meters per millisecond because radar occasionally sees the end of the channel (the edge of reflectivity) extending at these velocities. (Hewitt saw 10-25 m/ms, Mazur more than 50 m/ms, and Holmes more than 100 m/ms.) Radar could see but has not seen larger or smaller non-zero velocities, hence we tentatively conclude that over kilometer scale distances, lightning leader propagation into virgin air proceeds at 10 to 200 m/ms. ("Tentatively" because there are so few observations. "Kilometer scale" and "into virgin air" because echo fluctuations make it difficult to measure velocity over shorter distances or along still somewhat reflective channels.)

We do not know whether to imagine the isotherms as simple concentric cylinders, or as a turbulent shape of tongues and whorls. Nor, despite our attempts in appendix A1, does radar tell what diameter to imagine.

Most lightning echoes, instead of decaying immediately, remain relatively constant for tens or hundreds of milliseconds, then suddenly begin to decay at 0.2 - 5 dB/ms. (An example of relatively constant reflectivity followed by decay is shown in figure 1. But the decay rate shown here may be wrong, see caveats in the figure caption.) Two explanations have been offered for this behavior: Hewitt [1957] proposed that reflectivity remains constant for tens or hundreds of milliseconds because channels carry current for tens or hundreds of milliseconds; Mazur et al [1985] proposed that reflectivity remains constant for tens or hundreds of milliseconds because channels cool slowly from their return stroke temperature to an underdense plasma. The radar data is not presently able to rule out either explanation. We know from the interstroke field change that *some* channels are carrying current between strokes, but we cannot know that *all* channels with enduring reflectivity carry current.

Browne's echo, Krehbiel's K-change echoes, all of Cerni's echoes, some of Hewitt's echoes and some of our echoes (figure 6) began decaying immediately, at several dB/ms; these did not remain constant for tens or hundreds of milliseconds. We consider five possible explanations for this behavior; the channels might have been initially underdense, they might have cooled in milliseconds, the absorption just outside these channels might have rapidly increased, the electric field in these channels (and hence the collision frequency), might have rapidly increased, or these echoes might be from the short lived brush we believe to precede lightning streamers.

These channels were not initially underdense. Return strokes carry kiloamperes of current, and "underdense" requires so few electrons that the channel is incapable of carrying more than a few amperes.

(See appendix A2.) These channels did not cool to underdense in the first few milliseconds. Their echoes fall at 0.5 to 5 dB/ms, indicating a temperature decay of only 6 to 30 K/ms, and it is improbable that the initial temperature was by chance within tens of kelvin of the critical temperature. (Conversion from echo decay to temperature decay is discussed in figure 8.) Absorption did not rapidly increase an order of magnitude. Appendix A0 shows that 90% absorption requires either a temperature gradient less than 100 K/cm, or an electron diffusivity greater than 1 m^2/s. Neither is plausible. Rapidly rising electric field could not change the collision frequency more than a factor of two. Hence rapidly rising electric field could produce at most a factor of four decrease in the radar cross section of underdense plasma.

A last possibility is that rapid decays are not decay of the return stroke channel at all, but instead are decays of the brush we believe to exist at growing streamer tips. If so, this requires that the channel itself (which remains overdense until the usual tens or hundreds of milliseconds have passed) be too dim for the radars to see. This is possible; rapidly decaying echoes of Krehbiel, Cerni, and this study all began less than 10 dB above the noise, and Hewitt's rapidly decaying echoes were "weak". We tentatively suggest that rapidly decaying echoes come from the brush at the tip of lightning leaders.

We imagine a temperature decay of 3 to 30 K/ms because radar echoes decay at 0.2 to 5 dB/ms. Echo intensity is determined by the electron number density, the position of the channel in the beam pattern, the number of limbs in the beam, the radius of the channel, the roughness of the channel surface, the collision frequency, and the rate at which electron number density changes with radius. Of these, only electron number density can cause a 20 dB decay in a few milliseconds. Position in the beam is not changing rapidly in time. Decrease of number of channels in the beam cannot explain this decay, because observations in section 3 of this paper decay at the same rate as do other radar observations of lightning, and those observations are known to be of a single limb. Radar cross section varies slower than linearly with radius, and a factor of 100 change in radius is not plausible. A change in roughness or absorption might affect 3 cm measurements, but is negligible at VHF. The only remaining explanation for the 0.2 to 5 dB/ms decay of the radar echo is a decay of electron number density.

Lightning's radar reflectivity is independent of electron number density until that density falls below a critical density. Once below the critical density, the echo is (roughly) proportional to electron number density squared. (Details in appendix A2). So a 0.2 to 5 dB/ms decay in radar echo implies a 0.1 to 2.5 dB/ms decay in electron number density, which, from figure 8, implies a 7 to 150 K/ms decay of temperature.

Holmes et al [1980] and Mazur et al [1985] considered this slow temperature decay and inferred considerable ohmic heating. We doubted the existence of ohmic heating, because it required a surprisingly large electric field in the lightning channel. We expected our experiment to show that an isolated segment of the channel cooled faster than Holmes and Mazur thought. But we found cooling rates only an order of magnitude faster than Holmes and Mazur, and about the same as seen by Cerni and Krehbiel. And so we are forced to the same inference as Holmes and Mazur; there is significant ohmic heating in the 3700 K to 3900 K channels seen by radar.

Were ohmic heating and turbulent cooling both negligible, then the 7 to 150 K/ms cooling rate would determine channel radius. Uman and Voshall found cooling rates of 120 K/ms for a 2 cm diameter channel at 3500 K, and 30 K/ms for a 4 cm diameter channel at 3500 K. If we extrapolate, assuming that cooling rate is inversely proportional to radius squared, we get an 8 cm diameter for a channel cooling at 7 K/ms. So, if ohmic heating and turbulent cooling were negligible, the radar measurements of lightning would tell us that lightning's diameter varies from 2 to 8 cm.

8 cm exceeds any measured lightning diameter. Having already rejected all explanations other than a decay of centerline temperature for the observed echo decay, we must conclude that either cooling by thermal conduction is smaller than we thought, or lightning's diameter is larger than we thought, or ohmic heating is not negligible during the slowest decays. Thermal conductivity of air is very well known, and lightning's diameter probably known [Uman 1968], so we tentatively and reluctantly conclude that ohmic heating is not always negligible.

Our reluctance in this conclusion comes from a reluctance to believe that fields in the lightning arc are as large as ohmic heating requires. For a 4 cm diameter channel to cool at 7 K/ms instead of 30 K/ms requires 23 K/ms of ohmic heating; at the critical temperature for our radar (3900 K) 23 K/ms of Ohmic heating requires a field of 1.7 kV/m in the channel. At the critical temperature for Cerni's radar (3700 K) 23 K/ms

266

of ohmic heating requires a field of 2.7 K/ms. These fields are higher than measured in laboratory arcs [King 1961] or predicted for lightning arcs [Latham 1986].

If there is turbulent cooling, then the channel field (or radius) must be even larger. A turbulent diffusivity as small as 0.01 m^2/s causes cooling twenty times greater than that from conduction alone. In the absence of Ohmic heating, this would require channel diameters of 9 to 35 cm, instead of 2 to 8 cm. These diameters are unbelievably large; therefore, if there is turbulent diffusivity as large as 0.01 m^2/s, there must also be ohmic heating. The larger the field we are willing to imagine in the channel, the larger the allowed turbulent diffusivity. Diffusivity must produce cooling slightly greater than the heating; if we allow a field as large as 10 kV/m, we have 50 W/cm^3 (140 K/ms) of heating, which would be exactly cancelled by a diffusivity of (channel diameter^2 / 630 ms). If we imagine the field in the lightning channel as large as 50 kV/m, we get 900 W/cm^3 (2300 K/ms) which exactly cancels a diffusivity as large as (channel diameter^2 / 37 ms).

3. EXPERIMENT

Between strokes, is the channel to ground conductive or resistive? Ollendorf [1933] wanted a resistive channel to explain multiple discrete return strokes. Malan and Schonland [1951] wanted a resistive channel to explain electric field changes of different sign at different distances from the channel. But for some reason the resistive channel became unpopular. Brook et al [1962] wrote that "A small, steady 'dark' current would serve to maintain some minimum ionization in the lightning channel between strokes; otherwise, multiple strokes in identical channels would be rare." Ogawa and Brook [1969] found that Malan and Schonland's electric field changes could be explained without a resistive channel to ground, because interstroke charge transfers were more horizontal than vertical. Uman and Voshall [1968] found that the channel would remain conductive between strokes even without Brook's "dark current" if cooling were due to conduction alone (no turbulence) and if the channel radius were a few centimeters. Krehbiel et al [1979] restored the popularity of the resistive interstroke channel to ground by repeating Malan and Schonland's multiple station electric field measurements with many more stations. Some of his measurements are best explained by a charge transferred along the old lightning channel, but not all the way to ground, strongly suggesting a resistive channel to ground. After Krehbiel, Jurenka and Barreto [1985] still require a conducting channel to support the dart leader, while Uman [1987] allows that the dart leader is guided by either a conductive channel, or by a merely warm, non-conductive one.

So, between strokes, is the channel to ground conductive or resistive? In an attempt to unambiguously answer that question, we set up a 12.67 Ghz bistatic CW radar 26 meters from rocket triggered lightning at Kennedy Space Center. The 1.5 m wide, 1.5 meters high, and ten meters long volume observed (figure 3) included a 1.5 meter section of each triggered lightning channel (figure 4). We digitized the radar return every 300 microseconds. If the radar echo vanished between strokes, we would conclude that the plasma was underdense and therefore that the channel was resistive. We also hoped, with the same experiment, to measure the reflectivity of a known length of lightning channel at a known position and orientation in the radar beam, for comparison with theory. Finally, because we didn't realize how good were the existing measurements of echo decay rate, we wanted to make our own measurement of the cooling rate of the channel to ground.

The radar echo usually did fall to zero between strokes, but our equipment was inadequate to determine whether the echo always went to 0. (See figure 5) The 1.5 m segment of the lightning channel had a cross section between 210 cm^2 and 1700 cm^2. (See figure 5) For strokes followed by long continuing current, reflectivity remained relatively constant for tens or hundreds of milliseconds, then fell faster than 0.2 dB/ms. (See figures 5, 1) For strokes not followed by long continuing current, reflectivity began decaying immediately, at 3 to 5 dB/ms. (See figure 6) A tortuous 2.5 mm diameter wire hung in the same location occupied by lightning had a cross section of only 10 to 30 cm^2 / m. (See discussion, appendix A1) Ultraviolet photography did not reveal any dim channels perpendicular to the main stroke. (See figure 7)

Because the radar echo sometimes fell to our noise level between strokes (See figure 5), we conclude that the channel to ground sometimes becomes resistive between strokes. Because lightning reflected 100 times more than did a 2.5 mm diameter tortuous wire hung in the identical location, we conclude that the lightning channel diameter is at least greater than our 2.5 cm radar wavelength, perhaps as much as three times that large. (See Appendix A1) Because we saw only one channel, we conclude that our echo decay

267

was due to changes in that one channel, not due to reduction in the number of channels in the volume. Because our fastest decay rates (4 dB/ms) are in general agreement with the fastest decays seen in earlier observations, we tentatively conclude that previous observations of radar decay also describe changes in a single channel, not reduction of the number of channels in the radar resolution volume.

4. SUMMARY

Lightning is a sparse tree, with limbs hundreds of meters apart, not a volume filling thicket of channels tens of meters apart.

> The variation of echo extent with range fits one limb per resolution volume, and does not fit many limbs per volume. Thus the enduring limbs must be hundreds of meters apart. (Except perhaps at the tips, where Proctor's observation of echoes traveling in range suggest the existence of a small thicket of short lived channels only meters apart.)

Lightning channels are initially overdense

> A 4 cm diameter channel of marginally underdense plasma carries less than 3 A at 25 kV/m. Unless we have seriously underestimated lightning's diameter or electric field, current carrying channels must be overdense to radar.

Echo decay is mostly due to decrease in centerline temperature.

> Decay of some channels, absorption just outside the channel, or changes in channel radius cannot explain the observed decay.

Between strokes, the channel to ground becomes very resistive.

> The radar echo of the channel to ground drops 20 dB between strokes.
> Decay of some channels, absorption just outside the channel, or changes in channel radius cannot explain the observed decay, leaving a decrease in electron number density as the only plausible explanation. Underdense plasma is incapable of carrying lightning currents.

Centerline temperature decays 7-150 K/ms at 3700-3900 K

> Because echo decay is mostly due to centerline temperature decay, radar can measure temperature once the channel cools to an underdense plasma.

The electric field in 3700-3900 K channels is sometimes as high as several kilovolts per meter, unless the diameter of those channels is surprisingly large (> 10 cm) and the diffusivity surprisingly small (< 0.01 m^2/m).

5. SUGGESTIONS FOR FUTURE RADAR STUDIES OF LIGHTNING

A few measurements suggest that streamers have bushy heads, and that virgin propagation velocities range from 10 to 100 m/ms. Any confidence in these suggestions awaits additional measurements.

There is agreement that the decay rate varies from 0.2 to 4 dB/ms. Simultaneous measure of decay rate and channel radius would test *Uman and Voshall* [1968]. Simultaneous measure of decay rate and current would test *Latham* [1980].

Lightning mappers, able to see all space, not just a narrow beam, and potentially with high resolution in time, should make radars obsolete as a tool for measuring the position of lightning. A future experiment with both a mapper and a radar could test whether the mapper sees all the channels, and how accurately the mapper positions them. Lightning mappers often see propagation along old channels; the same experiment would determine whether these channels are cut off or still conducting just before such propagation.

APPENDIX A1. LIGHTNING'S RADIUS FROM ECHO MAGNITUDE

A few studies used uncalibrated radars; these studies can report the extent, velocity, duration, and decay rate of lightning echoes, but not the absolute magnitude of those echoes. Most studies used calibrated radars and quoted absolute cross sections in m^2. (See figures 9) Lightning's absolute cross section is the basis for

268

Dawson's claim that Browne's flash was overdense. Lightning's absolute cross section is the basis for claims by Holmes et al and Proctor claim that lightning is underdense. Lightning's absolute cross section is the basis for Williams et al's claim that lightning's mean channel length per unit volume is five kilometers per kilometer.

All of these claims require faith in some model of the lightning reflectivity. It has been difficult to test these models, because reflectivity depends on the position of lightning relative to beam center, on the length of channel in the beam, and on the lightning radius, none of which have usually been known. For our measurements described in section 3, we know the position of the channel relative to beam center, and the length of channel in the beam, though still not the channel radius. we now use a model of lightning reflectivity to compute the radius.

We saw seven events that for at least part of their duration were well centered in our 1.5 meter diameter beam. For four of these we transmitted and received E horizontal; we observed mean cross sections of 860, 1200, 1200, and 1700 cm^2. For two well centered flashes, (both on the same day) we transmitted and received E vertical; we observed mean cross sections of 200 and 600 cm^2. For one well centered flash, we transmitted E vertical and received E horizontal; we observed a mean radar cross section of 2 cm^2. (This may overestimate the true cross polar reflectivity; our antennas were loosely mounted, and could easily have become not 90 but 89 or 91 degrees out of phase.)

A 2.5 mm diameter tortuous wire, also well centered in the same 1.5 meter beam, produced a mean cross section of 22 cm^2. The lightning channel cannot have been more conductive than the wire; we may conclude with certainty that lightning's diameter exceeds 2.5 mm.

To know how much greater than 2.5 mm, we need a model for radar reflectivity of a tortuous conductor as a function of diameter. Williams et al provide such a model, valid so long as the diameter is much less than the radar wavelength. Their formula predicts a cross section per unit length of 16 cm^2 / m for a tortuous wire of 2 mm diameter. So in our 1.5 m tall beam the formula predicts 26 cm^2, and we saw 22 cm^2, agreement too good to be true. We tentatively conclude that the Williams' formula is at least approximately correct. The Williams formula does not predict a cross section as large as our lightning cross sections for any radius less than our 2.5 cm radar wavelength. We conclude that the overdense plasma had a radius greater than 2.5 cm.

Greater than our wavelength, the Williams et al formula is invalid; we need a new model. Our problem is that we don't know either the surface roughness or the precise tortuousity of the lightning channel. If we assume a perfectly smooth (mirror) cylinder, with angle of incidence equal angle of reflection, then geometrical optics predicts no reflectivity at all. (A beam transmitted up from the ground to a vertical cylinder is still going *up* from the ground on reflection, it is not scattered *down* to our receiver.) If we assume a perfectly rough (white) cylinder, with all power incident at any point isotropically re-radiated, then geometrical optics predicts radar cross section equal to half the geometrical cross section. (The other half is reflected in the other polarization.) The mirror surface approximation is not useful; we did see some reflectivity. The white surface approximation is not right either; transmitting E horizontal, we received more than a hundred times more power receiving E horizontal than when receiving E vertical. Lightning has some unknown intermediate roughness. The best we can do without knowing that roughness is assume that power is uniformly scattered, but that much less than half the incident power is rotated into the other polarization, and so predict a radar cross section equal to the geometrical cross section, at least when transmitting and receiving the same polarization.

A 210 to 1700 cm^2 geometrical cross sections in 1.5 meters imply 1.3 cm to 10 cm diameters. We trust previous measurements of lightning's diameter [Uman 1964] more than our "best we can do" model. 10 cm is almost certainly too large.

APPENDIX A2. OVERDENSE AND UNDERDENSE

"Overdense" means there are so many electrons that adding more causes almost no increase in radar reflectivity. "Underdense" means less electrons than that.

As lightning cools, electrons and ions recombine. At first, while the lightning is overdense, recombination doesn't change the radar echo. Once lightning becomes underdense, the radar echo begins to

decay. Radar sees when this decay begins and how rapidly it occurs. The onset of decay bounds possible temperature and current at that time. The echo decay rate determines the temperature decay rate.

The electron number density where decay begins is a function of radar frequency and collision frequency. Collision frequency is a function of radar frequency, electric field, temperature, and pressure. Electron number density is a function of electric field, temperature, and pressure (figure 8). If we know all but three of electron number density, radar frequency, collision frequency, electric field, temperature, and pressure, these three functions determine the three unknowns. Traditionally, radar frequency and pressure are known, electric field is guessed, leaving collision frequency, electron number density, and temperature.

For collision frequency, Hewitt used 20 GHz in no field, and 500 GHz in a field of 3 MV/m. Dawson used 61 GHz in no field and 370 GHz in 19 kV/m. Mazur and Doviak used 2.6 THz divided by the square root of the absolute temperature in kelvins. Williams et al use 45 GHz at 3500 K, and 37 GHz at 5000 K. All of these studies (except Dawson) used the collision cross sections for room temperature electrons, and assumed collision frequency to be directly proportional to electron velocity times density. Since the collision cross section for room temperature electrons is two and a half times smaller than the collision frequency for 3000 K electrons, all earlier studies (except Dawson) underestimated low field collision frequencies by more than a factor of two, and hence underestimated the critical electron number density by more than a factor of four.

Guessing the electric field causes at most a factor of three uncertainty in collision frequency, hence at most a factor of nine error in electron number density. The error is probably much less. Because electron number density varies an order of magnitude every 300 kelvins, (see figure 8) this causes at most a few hundred degrees error in the channel temperature.
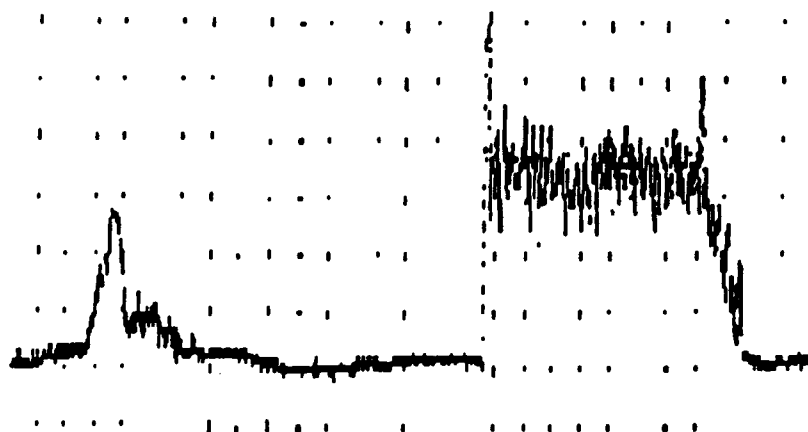
Figure 1. Radar cross section versus time. Our noise level is a cross section of less than one square centimeter. The first echo shown, with peak radar cross section of about 100 cm^2, is produced by the rocket passing through our beam. Although the rocket is connected to ground by a taut, 200 micron diameter copper wire, which remains in the beam, the radar cross section of this wire is apparently too small to be seen above the noise. One and a half seconds after the rocket's passage, the wire is vaporized by lightning, which, in this case, carried a single long continuing current to ground. This lightning flash remained well centered in the radar beam for its entire duration. This lightning had a mean radar cross section of 1200 cm^2. (Other flashes had cross sections of 200 cm^2, 600 cm^2, 860 cm^2, another 1200 cm^2, and 1700 cm^2.) The cross section remains roughly 1200 cm^2 for 700 ms, before beginning to decay. Unfortunately, once the decay begins, the data shown here for decay rate is meaningless, as our logarithmic amplifier does not correctly follow rapid changes in echo intensity. Some correct decays, measured for a different flash with a different amplifier, are shown in figure 6.
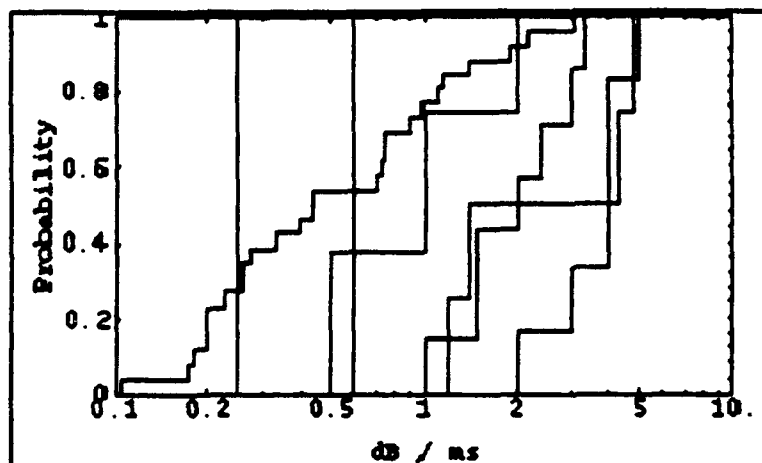
Figure 2. Summary of radar observations of the decay of lightning echoes.

decaydB/ms	wavelength	number of obs		beam volume
.1 - 3	10 cm	68	Holmes et al	500m x 250m x 150m
1.5	50 cm	5	Hewitt	2km? x 2km? x 300m
.25	10 cm	1	Mazur et al	350m x 350m x 37m
.6	50 cm	1	Proctor	2.8km x 2.8km x 600m
3 - 4.5	10.7 cm	3	Cerni	2.2km x 560m x 150m
.5 - 2	3	8	Krehbiel et al	? x ? x ?
1.2 - 4.8	2.4	4	This Study	1.5m x 1.5m x 10m

There is no apparent correlation between decay rates and wavelength. These are not two hump distributions, so there is no evidence for two distinct processes causing fast and slow decays.
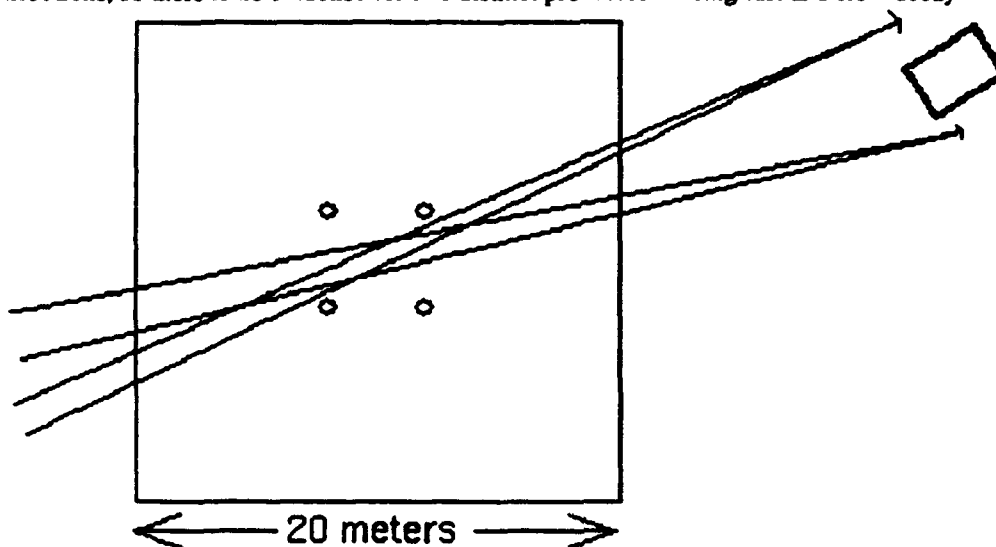
Figure 3. Top view of experimental setup. Antennas are mounted less than one meter off the ground; the beam intersection is about 8 meters off the ground. Any azimuthal misalignment of the antennas produced large noise levels from telephone poles in the beam. Any elevation misalignment produced negligible signal from lightning, as the beams then no longer intersect. For all of the data included here except figure 6, we have a sphere calibration on the same day, showing the antennas to be well aligned.

271
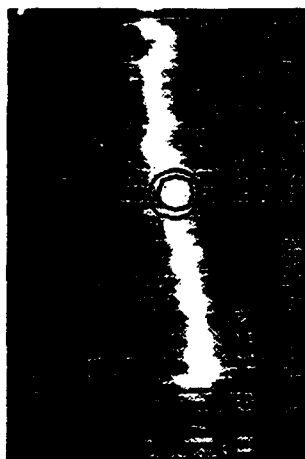
Figure 4. Experimental setup, as seen from shelter. This lightning flash produced the radar echo shown in figure 1. The 2 m lightning diameter shown here is presumably due to blooming in our CCD; photography shows a smaller radius. The 1.5 m diameter inner circle marks the boundary at which returned power from a calibration sphere is a factor of two lower than if the sphere is well centered. The 2.2 m diameter outer circle marks the boundary at which the returned power from a calibration sphere is a factor of four less than if well centered. (We measured echo from off-centered spheres as a function of distance off center on three separate days, with slightly different setups. The three different measured 3 dB and 6 dB diameters varied from 1.4 m to 1.6 m, and 2.1 m to 2.4 m, respectively.)
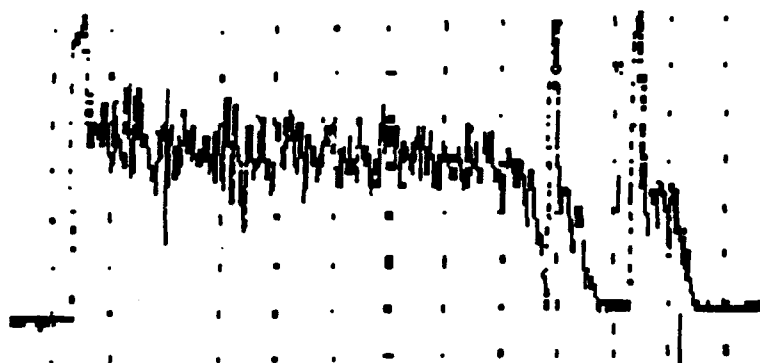
Figure 5. Running mean radar cross section for a lightning flash consisting of a long continuing current followed by three strokes. The radar cross section falls to the noise level at the end of the long continuing current, just before the first subsequent stroke. The radar cross section falls to the noise again for more than 50 ms between the first and second subsequent strokes. The radar cross section appears not fall to zero between the second and third strokes. Unfortunately, this amplifier does not follow rapid changes, so the rate of rise (not shown) and rate of fall (shown but *not correct*) of the echo are invalid. Mean cross section is expected to be correct, and the fact that the echo went to zero between the first and second subsequent strokes is expected to be correct. We cannot rule out the possibility that the cross section also went to zero between the second and third strokes, because we know that this amplifier responds to slowly to have seen this.

The lightning channel was (for flashes included in this paper) centered in the beam intersection volume, 20 m from the antennas. Voltage from the receivers was converted to radar cross section by comparison with the voltages from the same receivers as the radar observed metal calibration spheres of 7.5 cm to 30 cm diameter. These spheres were hung in the same location that the triggered lightning would occupy later in the day. If, for example, we report that a certain lightning flash had a cross section of 620 cm^2, we mean that flash reflected power halfway between the power reflected by our 510 cm^2 sphere and power reflected by our 730 cm^2 sphere.

The calibration was performed (usually with less than the full set of spheres) on 26 different days. The power reflected by a given sphere typically varied two dB between two measurements made the same day (usually the same hour), and varied four dB from day to day. Excluding cases where the antennas were obviously misaligned, the largest and smallest cross sections ever measured for a given sphere are less than 7 dB apart. We conclude that variations in radar performance and antenna alignment are responsible for less than 5 dB of uncertainty in our cross sections.

272
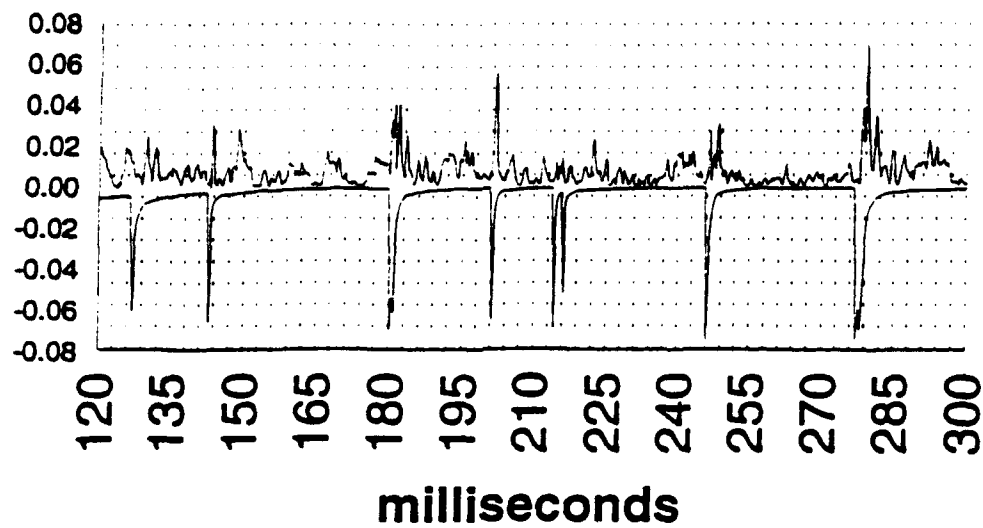
Figure 6. Instantaneous *relative* radar cross section and visible luminosity. The antennas were misaligned on this day; the cross section observed is a lower bound on the true cross section. This data comes from a rapidly responding detector, and so these decay rates are believed to be correct.
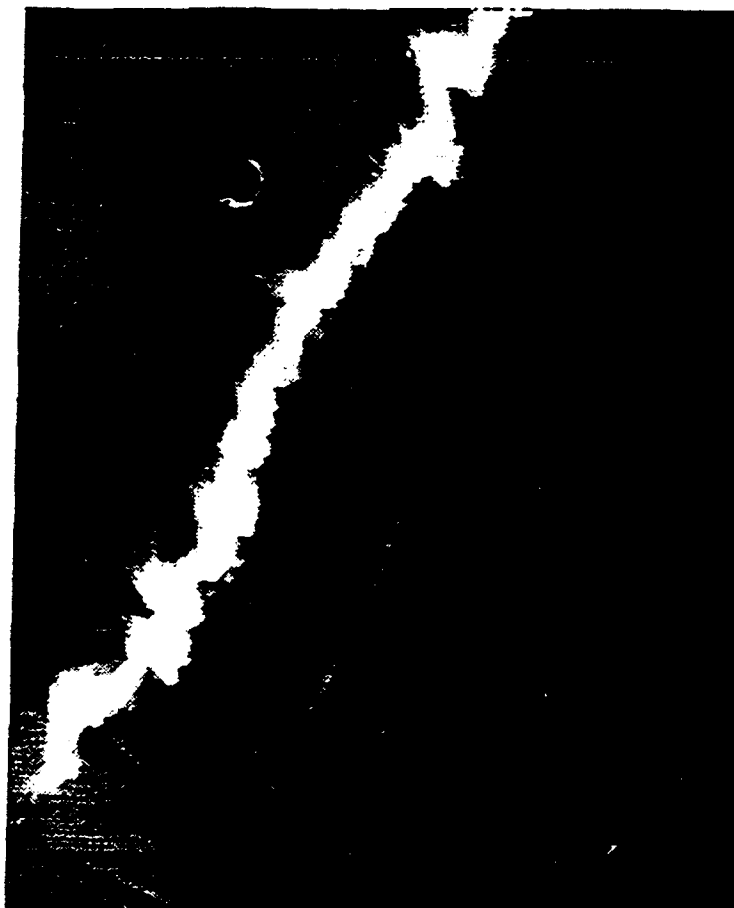
Figure 7. UV (295.9 + 2.5 nm) time exposure of lightning channel. The apparent breadth of the channel is likely due to channel motion during the time exposure.

273
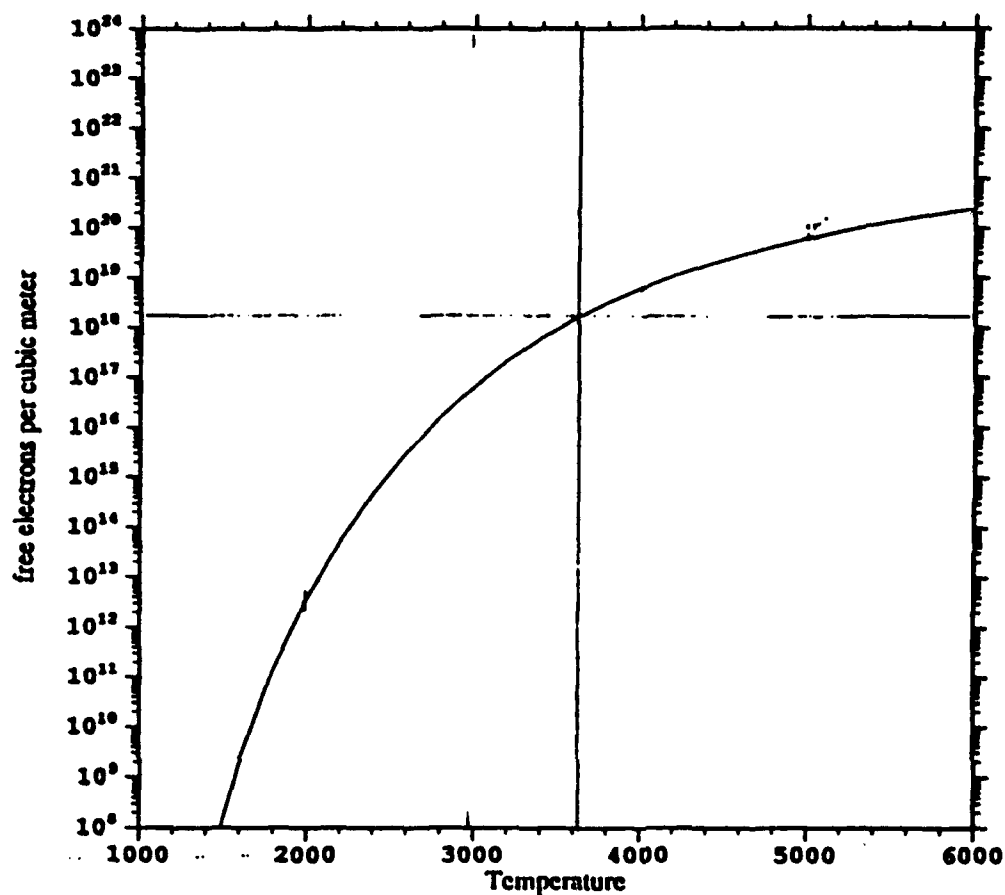
Figure 8. Thermodynamic equilibrium concentration of electrons in air. At these temperatures, the only electron donor of significance is NO+. All overdense plasmas reflect about equally, so radar cannot distinguish between temperatures larger than the critical temperature plotted here. Reflected power is proportional to electron number density squared, and electron number density falls rapidly with temperature, so radar cannot detect lightning at temperatures a few hundred degrees less that the critical temperature plotted here.

The critical electron number density is, except for the shortest radar wavelengths, and for electric fields exceeding 30 kV/m. At the shortest wavelengths, the critical density increases because radar frequency is no longer negligible compared to collision frequency. For electric fields exceeding 30 kV/m, the collision frequency increases as the electron temperature rises above the neutral temperature.

274
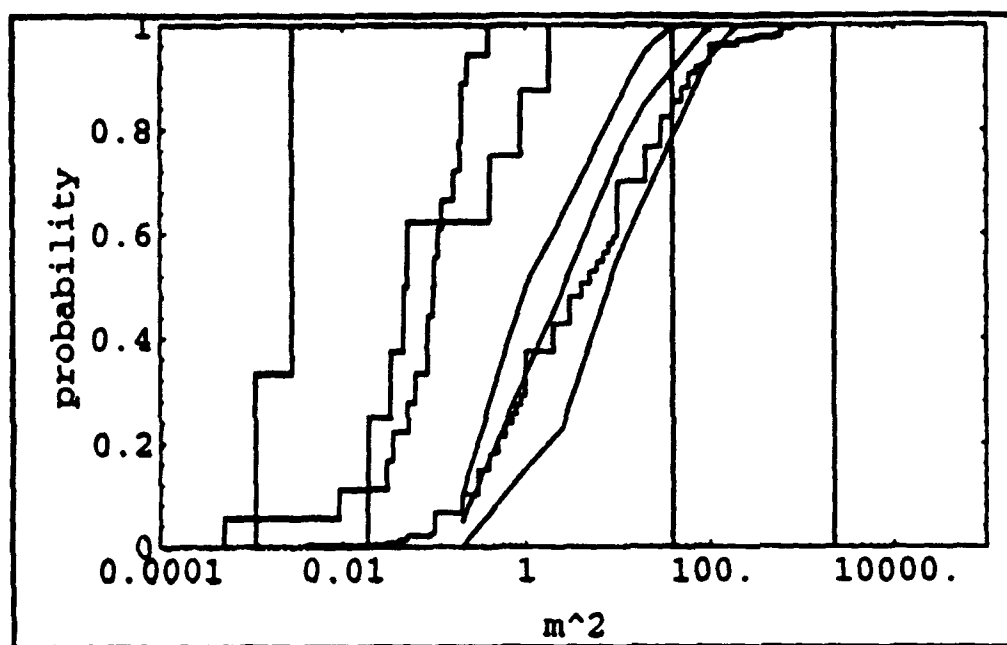
Figure 9. Lightning radar cross sections. The wavelengths and beam volume sizes at which these cross sections were observed are summarized in the following table.

cross section (m^2)	wavelength	number of observations		beam dimensions
2200	3.2	1	Browne	390 x 390 x 75
40	150	1	Pawsey	?
.0005 - .39	10.9	18	Holmes et al	420 x 210 x 150
.0001 - .003	10.7	3	Cerni	2200 x 560 x 150
.019 - 1.8	10.	8	Zrnic	340 x 340 x 150
.2 - 200	50	7	Proctor	2800 x 2800 x 600
.2 - 100	111	6	Proctor	2800 x 2800 x 30
.2 - 40	50	7	Proctor	2800 x 2800 x 30
.004 - 800	11	227	Williams et al	2200 x 2200 x 150
.2 - 1.7	2.4	6	This Study	1.5 x 1.5 x 10

REFERENCES

Atlas, 1958: Radar lightning echoes and atmospherics in vertical cross section. Recent Advances in Atmospheric Electricity, Pergamon Press.

Brook Kitagawa Workman, 1962: Quantitative study of strokes and continuing current in lightning discharges to ground, JGR, 67 649-659.

Browne, 1951: A radar echo from lightning. Nature, 167, 438.

Cerni, 1976: Experimental investigation of the radar cress section of cloud-to ground lightning. JAM, 15, 795-798.

Dawson, 1972: Radar as a diagnostic tool for lightning. JGR 77 4518-4527.

Hewitt, 1957: Radar echoes from interstroke processes in lightning. Proc. Phys. Soc. Lon. B70, 961-979.

Holmes, Szymanski, Szymanski, and Moore, 1980: Radar and acoustic study of lightning. JGR 85, 7517-7532.

Jurenka and Barreto 1985: Electron waves in the electrical breakdown of gases, with application to the dart leader in lightning, JGR, 90, 6219-6224.

Krehbiel, Rison McCrary Blackman and Brook, 1991: Dual-polarization radar observations of lightning echoes and precipitation alignment at 3 cm wavelength, Radar Met Conf. Paris 1991.

Krehbiel Brook McCrary 1979: An analysis of the charge structure of lightning discharges to ground, JGR, 84, 2432-2456.

Latham 1986: Anode column behavior of long vertical air arcs at atmospheric pressure, IEEE Trans. PS-14, 220-227.

Ligda, 1956: The radar observation of lightning. JATP, 9, 329-346.

Malan and Schonland 1951: The electrical processes in the intervals between the strokes of a lightning discharge Proc. Roy. Soc. 206A 145-163.

Mazur, Zrnic, and Rust, 1985: Lightning channel properties determined with a vertically pointing Doppler radar. JGR 90 6165-6174.

Mazur Rust and Gerlach, 1986: Evolution of lightning flash density and reflectivity structure in a multicell thunderstorm. JGR, 91, 8690-8700.

Ogawa and Brook 1969: Charge distribution in thunderstorm clouds, JGR, 95, 513-525

Ollendorf 1933: Versuch einer theorie der blitzsaule, Zeitschrift fur Physik, 27, 169-184.

Proctor, 1981: Radar observations of lightning. JGR, 86, 12109-12114.

Uman 1987: The lightning discharge, Academic Press, New York

Uman and Voshall 1968: Time interval between lightning strokes and the initiation of dart leaders, JGR, 68 497-506.

Williams Geotis and Bhattacharya, 1989: A radar study of the plasma and geometry of lightning, JAS 46 9 1173-1185.

Zrnic Rust and Taylor 1982: Doppler radar echoes of lightning and precipitation at vertical incidence. JGR 87, 7179-7191.

SECTION 2

Motivation:

    Triggered lightning begins with a positive leader. There exist laboratory experiments on positive leaders as long as 13 meters, and there exist theories producing excellent agreement with these experiments. If the first ten meters are the difficult part of triggering lightning, then these laboratory experiments should predict the aircraft potential and ambient field necessary to trigger lightning.

Conclusions:

    Data from the F106B, the CV580, and SPTVAR show no correlation between aircraft potential, ambient field, and lightning triggering. Data from the C160 shows aircraft potentials and ambient fields as predicted by positive leader theory.

What's in the rest of this section:

    Plots of aircraft potential and ambient field at the time of each lightning trigger on the F106B, CV580, and C160.

    A plot comparing NRL and ONERA measurements on the CV580

    A pictoral summary of leader physics known from laboratory studies of positive leaders.

For any given experimental program, pressure, humidity, aircraft velocity and aircraft capacitance don't change much, so we would expect the charge on the aircraft and the electric field around the aircraft to determine when the aircraft triggers lightning. Let's try to guess the relationship between ambient electric field and charge on the aircraft at the time the aircraft triggers lightning. If aircraft are making short positive leaders all the time, and the difficult part of triggering lightning is making a leader that can survive tens of meters from the aircraft, then there will be some critical ambient field at which the aircraft triggers lightning, independent of charge on the aircraft. (It's almost independent of the charge on the aircraft because, tens of meters from the aircraft, the field due to charge on the aircraft is small compared to the ambient field.) If the difficult part happens closer to the aircraft, then charge on the aircraft will be important, and less ambient field will be required if there is a large positive charge on the aircraft, and more ambient field if there is a large negative charge on the aircraft.

Indeed, one can quantitatively predict the number of volts per meter the ambient field can go down for each kV of aircraft potential. We assume that from an object as large as an aircraft, once positive streamers create a positive leader, continued development to a lightning discharge is inevitable. Laboratory streamers propagate until the potential drop to the electrode is less than 500 kV/m, and create a positive leader after a couple meters of propagation. For an aircraft body ten times longer than it is wide, positive charge on the aircraft increasing the aircraft potential by one volt allows a decrease of 1 volt / (1.5 aircraft length in meters) in ambient field. Negative charge on the aircraft requires an increase in the ambient field. A change in the aspect ratio from 10 to one to 100 to one changes the factor of 1.5 to a factor of 1.3. So we expect the fields and potential at time of lightning triggering to lie on a line of slope roughly 1.5 times the length of the aircraft. We shall see that only the C160 data look like this.

The NASA F106B was struck hundreds of times in 1982-1984, but apparently was not instrumented to determine aircraft charge and ambient electric field. Electric fields and aircraft charge just before 16 lightning strikes to the F106B in 1985 are reported by Rudolph et al 1989, and are plotted in the figure below.
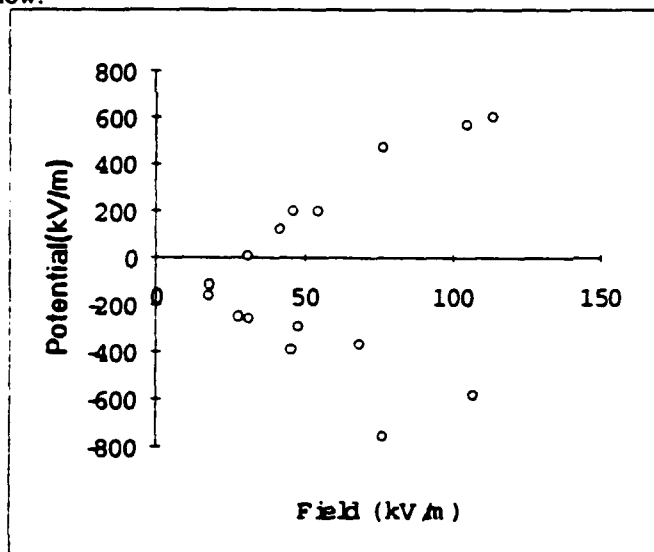
This does not show a single critical field for lightning triggering, instead it simply shows that the charge on the aircraft tended to be large when the aircraft was in a large electric field. This isn't what we expected; we expected that as the positive charge on the aircraft increased, the field required to trigger lightning would decrease. One possible explanation for this data is that triggering is caused by something other than the charge of the aircraft and the ambient electric field. A more likely explanation is that the mills on the aircraft were not carefully enough calibrated to distinguish between ambient fields and fields due to charge on the aircraft.

The NASA CV580 is the aircraft for which we have the best check of whether the fields and aircraft potential inferred are correct. The CV580 was simultaneously instrumented by ONERA and NRL, and the two teams separately analyzed data for 15 events they recorded together. The aircraft potential from the two measurements is almost uncorrelated; the ambient fields from the two measurements tend to differ in

direction by more than 60 degrees, and in amplitude by a factor of two. In any case, neither data set looks like what we expected, or like data from the F106B. See figures below.

NRL

Potential(KV) vs Ambient Field (KV/m)

ONERA

Potential(KV) vs Ambient Field (KV/m)

Ambient Field Comparison

NRL vs ONERA

Potential Comparison

NRL vs ONERA
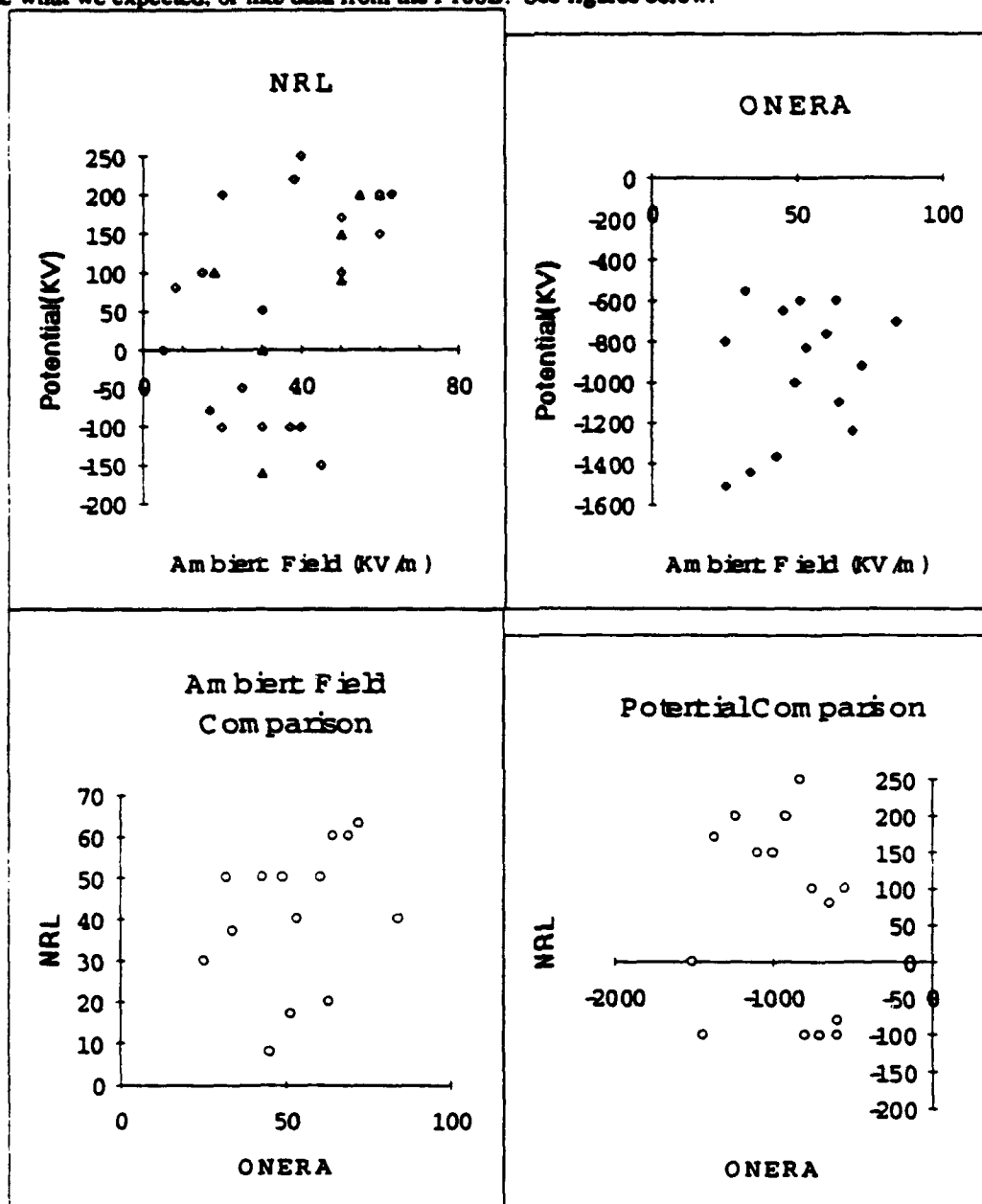
Two lightning strikes to SPTVAR in 1989 are reported in Jones et al 1990. The ambient fields were 5 and 7 kV/m, an order of magnitude smaller than fields reported on any other aircraft at time of triggering. The field at the mills due to the aircraft potential was "many times smaller" than that due to the ambient field. SPTVAR has been more carefully calibrated than any of the other aircraft, so this should not be explained as a bad measurement. SPTVAR cannot have triggered these flashes; laboratory streamers require an average field of 500 kV/m, and an ambient field of 40 kV/m is required to produce such a field for even one twentieth of the aircraft length. This suggests that SPTVAR was struck by lightning that began in some far away high field region, that SPTVAR did not trigger lightning.

Thirteen strikes to the Transall C160 are reported in Laroche et al, and plotted below. These appear to lie on two separate lines, each with the predicted slope. If instead of total field, we plot the larger of the horizontal field and 3/5 of the vertical field, the two lines coalesce, suggesting that one line corresponds to lightning initiation from the nose or tail, and the other to initiation from the vertical stabilizer. As the

279

analyzed total field decreases with increasing analyzed charge on the aircraft, it is unlikely that the slope is
due to the same pollution of the ambient field by the aircraft potential that characterized the F106B data.
The C160 data appear to be good.
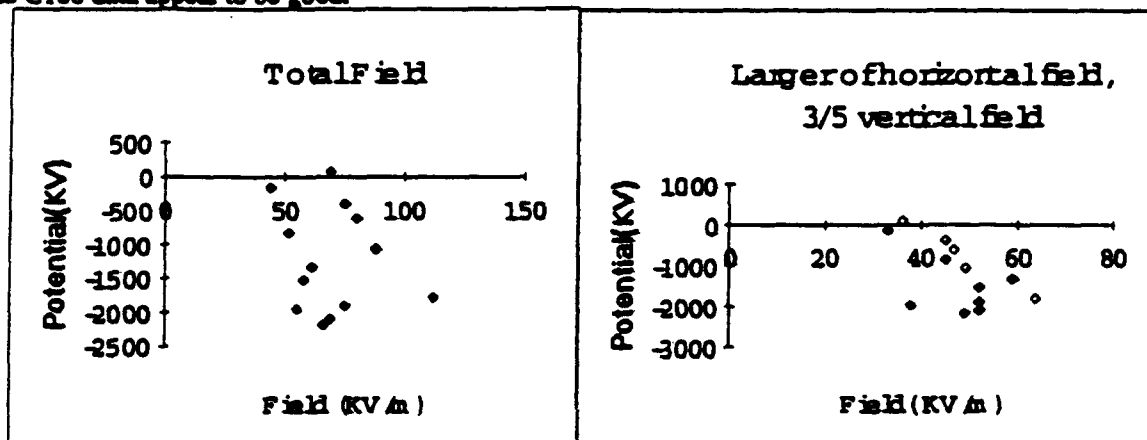
AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

INFRARED EMISSION OF GALACTIC HII REGIONS

Prepared by: Dr. Thomas Kuchar

Research Location: Phillips Laboratory,
                                      Geophysics Directorate
                                      Hanscom AFB, MA 01731

AF Research Colleague: Frank O. Clark

Contract Number: F19628-86-C-0224

1. Infrared Emission from HII Regions

This effort involved, in part, determining the infrared (IR) spectral properties of HII regions (areas of ionized gas surrounding young, hot stars) in the equatorial plane of the Galaxy. These properties were obtained from data and images observed by the Infrared Astronomy Satellite (*IRAS*). Some of the properties of the HII regions being examined are size, intensity, and color ratios and how these vary with position in the Galaxy.

The infrared emission for the set of 70 HII regions, with known kinematic distances, was examined using images from the *IRAS* Sky Survey Atlas in all 4 wavebands of observation (12, 25, 60, and 100 μm). These are radio bright ($f_{4860MHz} > 1$ Jy) HII regions located in the first Galactic quadrant: $30° \lesssim \ell \lesssim 60°, |\,b\,| \lesssim 0°.5$. The 'typical' HII region in this sample has a dust temperature $T(60/100) = 32$K (for λ^{-2} emissivity law), a diameter of 30 pc, and a luminosity ($\lambda \geq 50\mu$m) of $10^5 - 10^6 L_\odot$. However, there are several noticeable exceptions which are associated with the molecular ring and the W43 and W51 HII region complexes. These HII regions are hotter than average and appear brighter than average at all *IRAS* wavebands. It is interesting to note that the HII regions associated with the W49 complex have characteristics typical of the sample.

Initially, several IRAS images were examined over a 120 square degree region (8 fields that are 4 degrees on a side) in the first Galactic quadrant . The analysis of these images served as a prototype that will eventually cover all catalogued Galactic HII regions. Toward the end of this contract period the analysis was continued over four additional 12° square fields in the outer galaxy.

The spectral characteristics of these HII regions were examined using the low resolution spectrometer (LRS) data from *IRAS*. Polynomial fits were made to the red ($\lambda \gtrsim 12\mu$m) end of the spectra. The shapes of the spectra fell into two distinct classes based on the red end of the spectrum: flat or rising continua. Some of these spectra also showed silicate absorption at 10μm. After subtracting these fits from the spectra,

the residuals showed three known emission lines, [NeII], [NeIII], or [SIII], present in most spectra and possibility an unidentified emission line at $21\mu m$. Most of these lines are detected above the 3σ threshold, however some of them are only at the 2σ level.

All of the spectra are continuum dominated spectra with emission lines. No correlations were found among the emission lines in the LRS data and the spectra themselves. This included comparing the presence or absence of the three lines with the spectrum shape, linear slope, and 60/100, 25/100, and 25/60 flux ratios. One possible explanation for this lack of correlation is the source of the emission. The continuum in the spectra is mainly due to dust emission, whereas the emission lines arise from the ionized gas. The ionized gas is contained in a shell interior to the dust emitting region. Therefore the two sources of emission are not physically connected, although they both share the same heating source.

Cox (1990 Astr. & Ap., 236, L29) reported the presence of a $21\mu m$ band in the LRS data of H II regions. Cox argues that this is band emission by comparing the LRS with the point source catalog (PSC) fluxes at $25\mu m$. The $25\mu m$ bandpass covers some of the LRS wavelengths. In his sample of H II regions, the emission in the PSC fluxes underestimate the flux implied in the LRS data. As a result Cox argues that the feature seen at $21\mu m$ is not continuum emission, but rather band emission from molecules.

When the PSC fluxes are compared with the LRS data that I have analyzed, some of Cox's band features are evident. However, the feature is not as ubiquitous as Cox has claimed. Also, some of the LRS that show the same type of band features do not have the excess emission that Cox reports. Further investigation is needed in regards to this $21\mu m$ feature.

2. Rosette Nebula

The Rosette Molecular Cloud (RMC) is a site of recent (NGC 2244) and on going star formation (young stellar object AFGL 961). The RMC has properties typical of complexes that give rise to OB associations in the vicinity of the Sun. An early CO survey by Blitz & Thaddeus (1980 ApJ, 241, 676) shows that CO emission extends around almost the enitre periphery of the nebula. Later, Blitz & Stark (1986 ApJL, 300, L89) mapped the RMC in ^{13}CO using the 7m Bell Labs telescope. However, they did not include the area immediately surrounding the nebula. HI maps of the Rosette Molecular Complex show a shell of atomic gas that is expanding into the molecular cloud. High resolution IRAS images in all four bands show an unresolved ($< 1'$) compression front spatially coincident with this HI shell.

During this contract period, the Five College Radio Astronomy Observatory's array mapper, QUARRY, was used to map the expansion front in ^{13}CO to obtain high spectral resolution data of the molecular gas surrounding the nebula. These data will be used along with existing CO data to determine the kinematics and energetics of the molecular component of the expansion. The data are currently being analyzed.

3. Meetings

The following meetings were attended during the contract period. Posters or oral presentations are listed for their respective meetings.

Back to the Galaxy, University of Maryland, College Park, MD 12–14 October 1992: 'Infrared Emission of Galactic HII Regions' (poster)

American Astronomical Society 181st Meeting, University of Arizona, Phoenix, AZ 3–10 January 1993: 'The Boston University–Arecibo Galactic HI Survey' (oral presentation)

Celestial Backgrounds Review, University of Florida, Gainsville, FL 19–20 January 1993:

'IRAS Properties of H II Regions' (oral presentation)

American Astronomical Society 182nd Meeting, University of California, Berkeley, CA

5–10 June 1993: 'IRAS Properties of H II Regions' (poster)

4. Publications

The following papers were published or were in press during the contract period and are attached to this report:

Kuchar, T. A. 1992 **Back to the Galaxy**, *A. I. P. Conference Proceedings*, No. 278, ed. S. S. Holt and F. Verter, p 250: 'Infrared Emission of Galactic H II Region'

Kuchar, T. A. and Bania, T. M., 1993, *Astrophysical Journal*, Vol 414, 'A High Resolution H I Survey of the Rosette Nebula'

Kuchar, T. A. and Bania, T. M., 1993, *Astrophysical Journal*, submitted 'Kinematic Distances of Galactic H II Regions from H I Absorption Studies'

Infrared Emission of Galactic HII Regions

Thomas Kuchar
Phillips Laboratory/GPOB
Geophysics Directorate, Hanscom AFB, MA 01731

Email ID
kuchar@plh.af.mil

ABSTRACT

Infrared emission from HII regions has been extracted from the *IRAS* databases for a sample of Galactic HII regions identified by radio continuum and recombination line surveys. This investigation, based on a sample of radio HII regions, should lessen the bias towards local HII regions which are larger and brighter on average. The infrared luminosity, color and physical properties have been extracted from this data set as a function of Galactocentric distance. The global infrared properties of the HII regions are quite uniform with some very interesting exceptions.

1. INTRODUCTION

The general infrared properties of HII regions in the Galaxy were sought by compiling the IR fluxes listed in the *IRAS* Point Source and Small Scale Structure Catalogs. Also, the detailed properties for a group of $\sim$ 70 HII regions were studied using the *IRAS* Sky Survey Atlas for a 30 square degree field in the Galactic plane ($30° \lesssim \ell \lesssim 60°$). A list of 1000+ HII regions was compiled from the recombination line surveys listed below in the references. Distances (for $R_0 = 8.5$ kpc, and $\Theta_0 = 220$km s^{-1}) are derived from these references. Approximately 730 HII regions were listed in the PSC and 160 in the SSSC.

2. FLUX RATIOS FROM *IRAS* IMAGES AND PSC

Figure 1 shows the *IRAS* 12/100 and 60/100 flux ratios for a sample of HII regions located in the first Galactic quadrant: $30° \lesssim \ell \lesssim 60°, | b | < 1°$. The ratios are derived from the peak fluxes as they appear in images from the Sky Survey Atlas. Thus most of the HII regions in this sample are resolved in the images (*i.e.*, $\theta > 6'$). The histogram shows the mean ratio for the HII regions located in a 0.5 kpc wide bin. The error bars indicate the rms scatter about the mean. Both ratios show peaks near 6 – 6.5 kpc. These HII regions are quite possibly associated with giant molecular clouds and star forming complexes in the molecular ring.

The mean 12/100 ratio for the diffuse interstellar medium and molecular clouds is 0.06±0.02 (Laureijs 1989). Thus it appears that these HII regions are deficient in 12 μm emitters relative to 100 μm when compared with the interstellar medium. The mean 60/100 ratio for molecular clouds is 0.21±0.03. This is a smaller value the mean presented in Figure 1. However, this is not a surprising result. since the dust surrounding the HII regions is being be heated

Figure 1
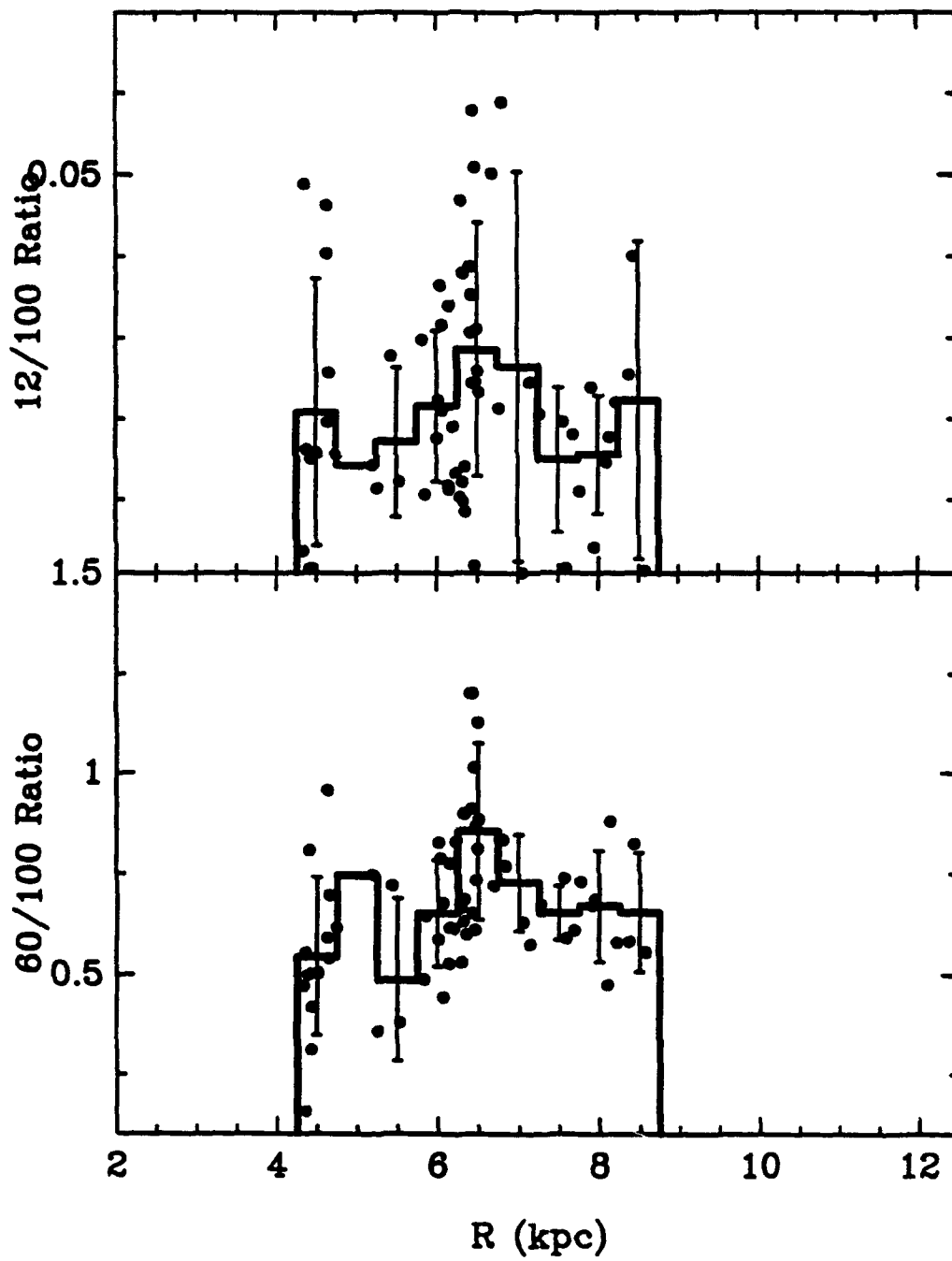
by the OB stars associated with the HII regions.

3. IR LUMINOSITIES

The far infrared color (60/100) temperatures and optical depths ($\tau \propto \lambda^{-2}$) were calculated for 48 HII regions in the Galactc plane for $30° \lesssim \ell \lesssim 60°$. Along with the the kinematic distances (Kuchar 1992), these data were used to determine the FIR luminosities of the HII regions. Figure 2 shows the distribution of luminosities in a face-on view of the Galaxy as seen from the north Galactic pole. The distribution peaks at Galactocentric radius of 6 kpc. The positions of the HII regions are marked by circles the size of which corresponds to the FIR luminosity of the HII region. Fiducial symbols mark the position of the Sun ($\odot$) and the Galactic center ($+$). Galactic longitudes are marked as well. The brightest HII regions ($L > L_\odot$) appear to be associated with the densest molecular clouds. This can be seen by comparing Figure 2 with a similar figure in Clemens *et al.* (1988) which shows the surface density of molecular gas.
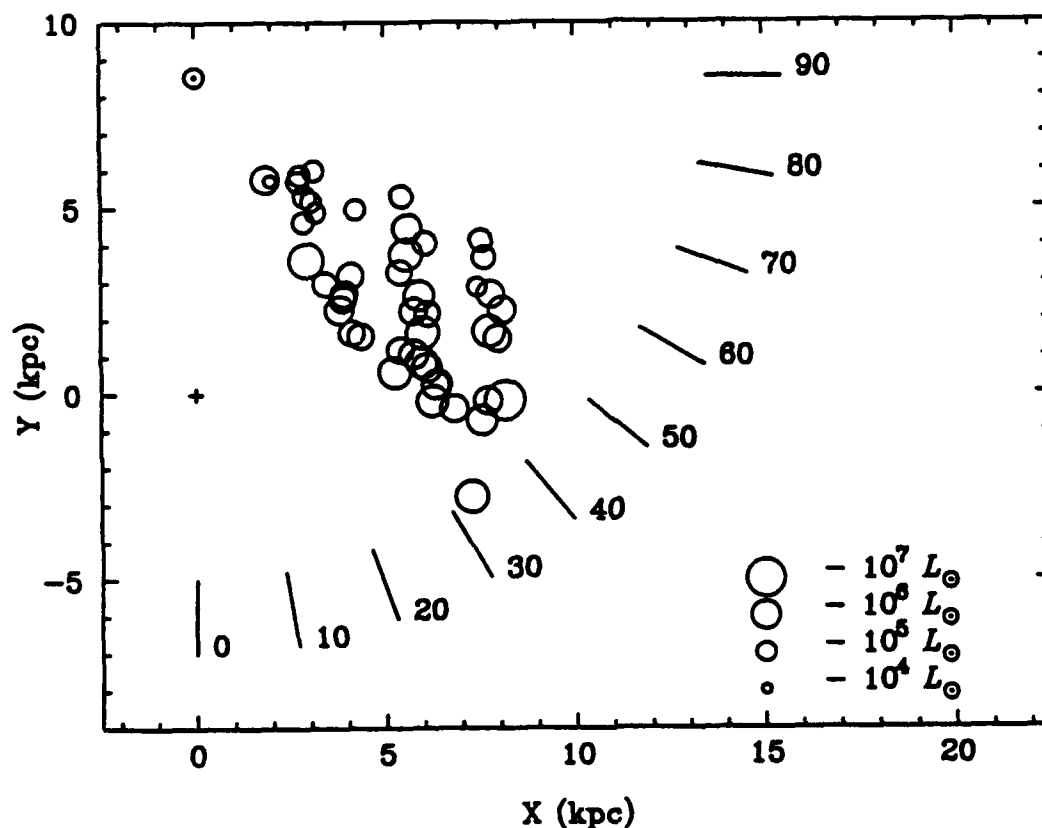
Figure 2

REFERENCES

Caswell, J. L. & R. F. Haynes 1987, A&A, 171, 261

Clemens, D. P., Sanders, D. B., & Scoville, N. Z. 1988 ApJ, 327, 139

Downes, D., Wilson, T. L., Bieging, J., & Wink, J. 1980, A&AS, 40, 379

Fich, M. & Blitz L. 1984, ApJ, 279, 125

Kuchar, T. A. 1992 Ph.D. dissertation, Boston University

Lockman, F. J. 1989, ApJS, 71, 469

Reifenstein, E. C., Wilson, T. L., Burke, B. F., Mezger, P. G., & Altenhoff, W.
    J. 1970, A&A, 4, 357

Wilson, T. L. 1980, in Radio Recombination Lines, ed. P. A. Shaver (Reidel:
    Dordrecht) p. 205

Wilson, T. L., Mezger, P. G., Gardner, F. F. & Milne, D. K. 1970, A&A, 6, 364

Wink, J. E., Altenhoff, W. J., & Mezger, P. G. 1982, A&A, 108, 227

A HIGH RESOLUTION H I SURVEY OF
THE ROSETTE NEBULA

Thomas A. Kuchar[1]

and

T. M. Bania

Department of Astronomy, Boston University
725 Commonwealth Avenue, Boston, MA 02215

Short Running Title: H I in Rosette Nebula

[1]current postal address: Phillips Laboratory/GPOB, Geophysics Directorate, 29 Randolph Road, Hanscom AFB, MA 01731

ABSTRACT

As part of the Boston University–Arecibo Galactic H I Survey, we analyzed the gross physical properties of the atomic gas associated with the Rosette Molecular Complex (RMC) which was derived from a new, high resolution 21cm survey. At the distance of the RMC, the 4′ angular resolution of the survey corresponds to a linear size of 2pc. The H I maps presented here cover an area of 110×60 pc. If the 21cm emission is optically thin, then the survey mapped an H I mass of $1.5 \pm 0.1 \times 10^5$ $M_{\odot}$. The optical nebula is surrounded by three, dense H I regions, which roughly trace out the boundary of an H I shell. The shell is located at the periphery of the atomic cloud and is expanding into the cloud at $4.5\,\mathrm{km\,s^{-1}}$. The expansion center is located at $(\ell, b, v) = (206.^{\circ}266, -2.^{\circ}083, 16.0\,\mathrm{km\,s^{-1}})$. A constant expansion velocity implies a dynamical age for the shell of approximately 4 million years, which is consistent with the age of the central cluster NGC 2244. The total kinetic energy of the expansion is estimated to be 3.8×10^{48} ergs. This represents about 2% of the total (radiative + mechanical) energy available from the central cluster. Thus the central cluster can provide sufficient energy to power the flow through a combination of stellar winds and radiation.

Subject headings: ISM: clouds – ISM: individual (Rosette Nebula) – radio lines: ISM

1. Introduction

The Rosette Molecular Cloud (RMC) has properties typical of complexes that give rise to OB associations in the vicinity of the Sun. Blitz & Thaddeus (1980) estimated the total mass of molecular hydrogen in the complex (over an area of 2.51 square degrees) to be 1.3×10^5 $M_\odot$ with a comparable mass in atomic hydrogen. The RMC is also a site of recent star formation. The central cluster in the Rosette Nebula, NGC 2244, contains several main sequence O stars (Perez et al. 1987). The presence of an O4v star indicates that the age of the cluster is $\lesssim 5$ Myr (Cox & Guili 1968) but may be as young as 10^5 yrs. In fact there are indications of active star formation in the RMC. The complex contains a number of infrared sources, many of which are associated with molecular clumps (Cox et al. 1990). The most luminous of these IR sources, AFGL 961, is associated with a molecular outflow and thus may be the result of an embedded, young stellar object (Lada & Gautier 1982).

The Rosette complex has also been investigated in several wavelength bands. The dust and infrared properties were examined by Celnik (1986) and Cox et al. (1990). The properties of the molecular gas have been discussed by Blitz & Thaddeus (1980) and more recently by Blitz & Stark (1986) and Blitz (1987, 1991). The ionized gas of the H II region of the Rosette Nebula has been investigated by Celnik (1983, 1985). Raimond (1966) conducted a coarse H I survey with a spatial resolution of $0°56$ and a velocity resolution of 2.1 km s^{-1}. He mapped 22 square degrees of the complex and determined an H I mass of 2.0×10^5 $M_\odot$ (for an assumed distance of 1600 pc). However the properties of the atomic gas in the complex have yet to be investigated in detail.

In this paper the results of a high resolution, neutral atomic hydrogen survey of 8.4 square degrees of the Rosette complex are presented. In the following sections the density and velocity structure of the Rosette complex are discussed including the effects of the cluster on the surrounding gas. Throughout this paper we assume that the distance to the RMC is 1600 pc, which is the distance to NGC 2244 (see review by Perez et al. 1987 for the distance determination and associated errors).

2. Observations and Data Reduction

The HI observations were made over a period of a few months in the spring of 1988 (Epoch 3 of the Boston University–Arecibo Galactic HI Survey; see Bania & Kuchar 1993) with some additional data acquired in October 1989 with the 305m Arecibo telescope[1].

---

[1]Arecibo Observatory is part of the National Astronomy and Ionosphere Center (NAIC) which is operated by Cornell University under contract with the National Science Foundation.

The data were observed with the 21cm flat feed in frequency switched mode with the reference frequency shifted 2.5 MHz from the HI line. The feed has a center frequency of 1407 MHz and a bandwidth of $\sim$ 40 MHz. The mapped region covers a rectangular grid in (ℓ, b) space and is sampled at 4', the HPBW of the flat feed. The data cover approximately eight square degrees in the range $204°866 \leq \ell \leq 208°799$, $-3°150 \leq b \leq -1°016$ (see Table 1 for the observational parameters) and have a velocity resolution of $0.52 \, \mathrm{km \, s^{-1}}$. At a distance of 1600 pc, then approximately 110×60 pc of the Rosette complex has been mapped with a spatial resolution of 2 pc.

All of the spectra had linear baselines removed and were corrected for feed gain variations with zenith distance. The HI data presented below and in the tables are expressed in terms of brightness temperatures, whereas the figures are expressed in antenna temperatures. The conversion between the two temperature scales is $T_b = 1.3 T_A$ (cf. Bania and Kuchar 1993 and Bania and Lockman 1993). Once the gain correction was applied to all the data, secondary calibration effects were evident. The integrated intensity maps showed a 'mosaic' pattern, i.e. portions of the maps showed pronounced, rectangular-shaped jumps in the intensities. The edges of these mosaic patterns outline the observations that were made on a specific day. These daily variations ($\sim$ 5%) in the feed sensitivity can be caused by changes in the weather (e.g. rain running down the antenna feeds changes the feed sensitivity).

Daily correction factors were derived according to the procedure described below.

The correction factors for a specific date have an uncertainty of at most 1%; however, a typical uncertainty is $\sim 0.5\%$. The resulting set of correction factors ranged from 0 to 8%, with the average correction being just less than 3%. The daily corrections were applied to the entire H I data set. The resulting integrated intensity maps showed a considerably reduced mosaic effect.

Changes in the receiver system temperature from the expected value were used to correct the intensity calibration of the data set. If the feed sensitivity (and hence the gain as well) changes, this will be reflected in the measured system temperature, T_{sys}. Thus, an anomalous increase in T_{sys} can be an indication of a change in the feed sensitivity. If the gain varies as a result of daily changes in the feed sensitivity, then we can attempt to model this variation as a change in the expected T_{sys}. The empirical T_{sys} curve for the BU-Arecibo Survey was used to estimate the expected T_{sys} for the Rosette data. The Survey curve was determined from observations near the galactic plane ($b \sim 3°$) devoid of strong continuum sources. It was also calculated using the off-frequency ($\nu = 1417.5$ MHz) system temperature so that there is no contribution from Galactic 21cm line emission. Therefore comparing the off-frequency system temperatures, $T_{sys,off}$, of the Rosette data with the empirical Survey T_{sys} curve gave an indication of the correction needed to compensate for the daily variations.

The Rosette mapping procedures assigned scan numbers each day to all the spectra that filled in a particular map section. For a given day's observations, the difference between $T_{sys,off}$ for all spectra with the same scan number and the Survey T_{sys} curve was computed. A correction factor, k, was determined to minimize this difference such that the average difference for a particular scan number was less than 0.1 K:

$$< T_{sys,off} - k \times T_{sys} > \leq 0.1 \, \text{K}. \tag{1}$$

Observations toward continuum sources were excluded from the calculation. Most days had observations with more than one scan number. The correction factors for all scan numbers observed on the same day were averaged together. This average then became the correction for that date.

Finally, the entire map was smoothed in angle by convolving the original data with a circular, two-dimensional Gaussian with a FWHM of $4'$, the beamwidth of the flat feed. The convolution was performed over ~ 1.25 beamwidths (3σ of the Gaussian), thus the

smoothing does not seriously compromise the spatial resolution of the data. The overall result is to reduce the noise of the individual spectra by a factor of 3 to ~ 0.15 K and to remove artificial spatial variations that appear smaller than the beamwidth. *This smoothed data set will be used in all further analyses.*

2.1 Density Structure

A rough estimate of the ambient H I density at the outer boundary of the Rosette Nebula can be made by using Hα emission to delineate the edge of the Strömgren sphere:

$$n = \left[\left(\frac{3}{4\pi} \right) \left(\frac{S_*}{\alpha(T_e) R_s^3} \right) \right]^{\frac{1}{2}} \quad [\text{cm}^{-3}] \tag{2}$$

where n is the total (neutral and ionized) H density, S_* the number of ionizing photons, T_e the electron temperature, and R_s is the Strömgren radius. The recombination coefficient, α, is taken from Mathews & O'Dell (1969). Here we assume that ionization of H I is precisely balanced by recombination at R_s. Panagia (1973) gives the number of ionizing photons (above the Lyman limit) for NGC 2244 as 1.44×10^{50} s^{-1}. The (non-LTE) electron temperature for the Rosette was determined by Celnik (1985) to be 5800±700 K.

The Strömgren radius was inferred from Celnik's (1983) Hα survey of the Rosette Nebula. He has calculated the mean flux density of Hα emission in rings of width 2′, centered upon the central minimum at $(\ell, b) = (206°247, -2°106)$. The radius of the Hα emission out to the noise limit of the data is approximately 68′. The resolution of the data is 4′75, which is taken as the error in this measurement. For an assumed distance of 1600 pc, the edge of the Hα emission is 32 ± 2 pc from the center. If this is taken as R_s, equation (2) gives an ambient H density of 10 ± 2 cm^{-3}.

The radius of the nebula as it appears on the red Palomar Observatory Sky Survey plate is ~ 38′ (= 17.7 pc). If this is taken as the Strömgren radius, then the corresponding density is ~ 60 cm^{-3}. Thus the H I density in the vicinity of the nebula is uncertain by nearly an order of magnitude based on these crude estimates. However, since the medium is not composed purely of hydrogen, other elements are likely to be ionized and will also contribute to n_e.

For a homogeneous and isothermal medium, the H I column density is given by:

$$N(H\,I) = 1.8 \times 10^{18} \int T(v) \left[\frac{\tau(v)}{(1 - e^{-\tau(v)})} \right] dv \quad [\text{cm}^{-2}]. \qquad (3a)$$

where $T(v)$ is the brightness temperature at LSR (local standard of rest) velocity v. For optically thin emission, the above expression can be simplified:

$$N(H\,I) = 1.8 \times 10^{18} \int T(v) dv \quad [\text{cm}^{-2}] \propto T_s \tau \Delta v. \qquad (3b)$$

Thus a simple map of velocity integrated H I intensity can be regarded as a column density map provided that these assumptions hold.

Figure 1 shows the integrated intensity map for the smoothed H I data. The map was integrated over velocities 3.1 to 24.2 $\text{km}\,\text{s}^{-1}$, which correspond to the velocities of CO emission associated with the RMC based on the Bell Labs CO Survey (Blitz and Stark 1986). Also these velocities fall within the velocity range that Raimond (1966) determined for the H I gas in the area of the complex. His cloud C, which is coincident with the Rosette Nebula, has a velocity range of 0 to 25 $\text{km}\,\text{s}^{-1}$.

The contours in the figure are expressed in $\text{K} \cdot \text{km}\,\text{s}^{-1}$, but can also be considered contours of column density using eq. (3b). There are several striking features in this figure. The position of the optical nebula is evident in this map as a local minimum at $(\ell, b) \simeq$ (206°2, −2°1). Below, and to the left of this position are regions of enhanced column density. However, there is a 'valley', a long, narrow region of decreased H I column density (or H I emission), that separates these two regions of higher column density. This valley extends to the position of the optical nebula. To the upper right of the nebula, the H I column density decreases. The left edge of the figure shows a rather smooth distribution of H I emission, in contrast to the area that surrounds the nebula.

Assuming that the H I emission is optically thin, the average column density over the mapped region is $2.7 \pm 0.2 \times 10^{21}$ cm^{-2}. The total H I mass ($= m_H N(H\,I) \times$ surface area) in the mapped region is approximately $1.5 \pm 0.1 \times 10^5 M_\odot$. This is comparable to the H I mass estimate made by Raimond (1966), where he calculated a mass of 2×10^5 $M_\odot$ over an area of 16 square degrees. Using the volume density calculated above and the average column density, N, we estimate the line of sight depth of the complex to be L $\sim$ (N/n) = 87 pc for n = 10 cm^{-3}. This depth is comparable to the diameter of the Hα emission, 64

pc. If the second estimate of the volume density (60 cm^{-3}) is used, this implies a depth L ~ 14 pc.

It is interesting to note that the absolute minimum of the H I emission in Figure 1 does not occur near the center of the Rosette Nebula but rather toward the upper right corner of the figure. The contours here show a relatively steep rise in the integrated H I intensities suggesting that this area marks the periphery of the H I cloud associated with the RMC. In Raimond's (1966) survey of the region, a portion of the half-peak column density contour of his 'C' cloud coincides with this area of low column density. Thus the 'C' cloud contour may be regarded as the edge of the RMC's atomic component.

The maximum H I column density occurs south of the optical nebula at $(\ell, b) = (205°931, -3°017)$ with a value of $3.4 \times 10^{21} \text{ cm}^{-2}$. The mass of H I contained within an area roughly 20′ in radius about this point is approximately 7000 $M_\odot$. Also, there are two other features of high column density located just outside the nebula at $(\ell, b) = (206°865, -1°483)$ and $(207°199, -1°683)$. Each of these is roughly circular with a radius of 12′. The mass of H I contained in these two regions is approximately 2500 and 2100 $M_\odot$, respectively. These three density enhancements comprise 10% of the H I mass of the mapped area.

An expanding shell surrounding the optical nebula has been detected in the H I data. The three density enhancements roughly trace the outer boundary of the shell. The shell is irregularly shaped and appears on the periphery of the H I cloud. Raimond discusses the evidence in his data for a thin H I shell surrounding the nebula. He derived a radius of 1° (25 pc) and a mass of 8.5×10^3 $M_\odot$ for the shell. However, the resolution of his data was too coarse (in space and velocity) to be certain of these parameters. The present data have sufficient resolution to determine the velocity structure in the shell. These characteristics are derived in the next section.

2.3. Velocity Structure in the Nebula

The presence of OB stars in NGC 2244 should affect the kinematics of the gas in their immediate environment. The radiation pressure and the stellar winds from the cluster members should accelerate the material surrounding them. In the ideal case of a

uniform medium, a centrally located cluster will cause the gas to expand spherically and create a shell of material (Castor et al. 1977). Thus expansion of the gas should be evident in the H I map. As the nebula expands, the H I gas and dust would be swept up by the shell resulting in a decrease of the H I density towards the center of the nebula. If the H I shell is optically thin, isothermal, and of uniform density, then changes in the brightness temperature, T_b, will trace the pathlength through the shell traversed by the line of sight:

$$T_b(v) = T_s(1 - e^{-\tau(v)}) \simeq T_s \tau(v), \tag{4}$$

where τ is the optical depth at velocity v and T_s is the H I spin temperature. Thus as the line of sight crosses longer path lengths through the shell, the optical depth increases and, as a result, T_b increases.

This expansion would be apparent in a longitude–velocity diagram of the data observed toward the center of the expansion. The shell expands in three dimensions, but an (ℓ, v) diagram shows only the line of sight component. A line of sight that passes through the expansion center will show the systemic velocity, V_s, of the complex along with the full component of the expansion velocity, i.e. $V_s \pm V_{exp}$. As the line of sight moves away from the expansion center, the projected component of V_{exp} along the line of sight decreases. At the edge of the shell, the expansion is tangent to the the line of sight and the only velocity component observed is V_s. (The position at which this occurs can be used to define the radius of the shell.) The result is an expanding shell that shows up in velocity space as a circular feature centered at the velocity of the nebula. Below we illustrate this by modelling the emission that is expected from a spherically expanding, uniform shell.

The model shell is taken to be at a uniform temperature and density and is expanding spherically at a constant velocity. Positions along the line of sight are divided into a total of N cells. The gas is confined to the cells within a radius $R_s \pm \Delta R/2$ of the expansion center, where ΔR is the shell thickness. This gas is assumed to have an internal velocity dispersion which is incorporated in the optical depth profile. For simplicity, the optical depth profile is taken to be Gaussian, with the optical depth at velocity v of the ith emission component being:

$$\tau_i(v) = \tau_0 \exp\left(-\frac{1}{2}\left(\frac{v - v_i}{\sigma_v}\right)^2\right), \tag{5}$$

where σ_v is the velocity dispersion, v_i is the velocity of line center, and τ_0 is the optical depth at line center (assumed to be constant for a uniform shell). A consequence of this line shape is that the optical depth profiles for different cells at different velocities, v_i, will overlap in velocity space. Thus the observed brightness temperature at a particular LSR velocity will be the sum of all the emission at that velocity:

$$T_{b,n}(v) = \sum_{i \neq n}^{N} T_{b,i}(v) + T_s \tau_n(v) \exp\left(-\sum_{i=1}^{n-1} \tau_i(v)\right). \tag{6}$$

Here, the emission from the nth cell at a velocity v is the sum of the emission at the same velocity for the other $N-1$ cells along the line of sight. Also the emission from the nth cell is attenuated by the material in the $n-1$ foreground cells. An (ℓ, v) diagram of an expanding shell was computed using the above model and the parameters listed in Table 2. Figure 2 shows an (ℓ, v) diagram for this model taken through the expansion center. The figure includes only emission from the shell, no quiescent cloud emission is included in this model.

The parameters listed in Table 2 were chosen to reproduce the gross properties of the H I shell in the Rosette Nebula (e.g. the size, velocity, and column density [$\sim T_s \tau_0 \sigma_v$] of the shell). The average density ($\sim N(H I)/\Delta R$) of the shell using these parameters is 10 cm^{-3}, which was set from the Strömgren sphere calculation in §2.2. The key feature in Figure 2 is the limb brightening that occurs at the edges of the shell where the path length through the shell and thus the optical depth is a maximum. The H I optical depth peaks at 0.17, given the model parameters. Also, the contours at the edges of the shell show a steep rise in T_b with longitude, whereas the other portions of the shell do not.

Observing limb brightening is a necessary condition for identifying an optically thin, expanding shell in the H I (ℓ, v) diagrams. Moreover, the neutral gas density is expected to decrease toward the center of the shell. This is due not only to the expansion of the material inside the shell but also to the degree of ionization as well. If the gas interior to the shell is isothermal, then this will show up as a decrease in T_b. It should be emphasized here that all of these conclusions are based on an ideal case. The contour levels in Figure 2 were chosen to enhance the effects of limb brightening. The clumpy nature of the RMC will necessitate modifying some of these assumptions. However, the gross properties described above should be observable in the H I data.

Figure 3 shows an (ℓ, v) diagram derived from the observed H I data at $b = -2°02$. Apparent in this figure is an oval shaped structure centered at $(\ell, v) \simeq (206°27, 16.0\,\mathrm{km\,s^{-1}})$. The limb brightening predicted by the model is apparent in this figure and several other similar images. This is emphasized in Figure 4 which shows the model superimposed over the H I data. This leads us to identify this structure as an expanding shell. From examining (ℓ, v) diagrams at other latitudes, the shell extends $\sim \pm 4.5\,\mathrm{km\,s^{-1}}$ in velocity. The maximum angular size of the shell from the expansion center to its interior edge is $0°64$. The thickness of the shell is somewhat more difficult to determine because of both the clumpy nature of the RMC and the departure from spherical symmetry. From examining velocity-integrated maps such as Figures 1 and 5, the shell thickness ranges from approximately $8'$ to $20'$. The expansion center was determined to be $(\ell, b, v) = (206°266, -2°083, 16.0\,\mathrm{km\,s^{-1}})$. This position is very close to (within $1.'8$, or 0.8 pc) the $H\alpha$ minimum. Since this is closer than one H I pixel, the positions are essentially coincident. The fact that these two centers are coincident in position is taken as a further indication that the structure seen in the H I data is a result of the expansion of the nebula. Also, the expansion center is $3.'3$ (1.5 pc) from the cluster center (defined using the OB stars listed in Cox *et al.* 1990).

Figure 3 also shows that a segment of the shell is almost absent at higher velocities ($206°2 \lesssim \ell \lesssim 206°4$, $V \gtrsim 20$ km s^{-1}). If H I emission at these velocities is from the far edge of the shell, then the shell may have expanded beyond the periphery of the H I cloud and 'punched' a hole through the back of the cloud. If the expansion of the shell has carried gas beyond the cloud boundary, then the radius of the shell can be used to determine the depth of the H I cloud. The distance from the expansion center to the back edge of the cloud can be taken to be somewhat less than the radius of the shell. A shell radius of $0°64$ corresponds to a linear size of 17.9 pc. Twice the radius is consistent with the depth of the cloud derived in §2.2 if the mean density is 25 cm^{-3}. The size is dependent on the shell being roughly spherical.

An age for the shell can be estimated if the expansion velocity is assumed to be constant over the life of the shell. For $V_{exp} = 4.5$ km s^{-1}, the dynamical age of the shell is $\lesssim 4.0$ Myr. Ogura & Ishida (1981) estimate an age of 3 to 5 Myr for NGC 2244 from fitting isochrones to the ZAMS turn–off point of the brightest stars. Thus the expansion age is consistent with that of the central cluster. The LSR velocity of the expansion center

$(16.0 \pm 0.5 \ \mathrm{km\,s^{-1}})$ is essentially that of the $16.7\pm1.3 \ \mathrm{km\,s^{-1}}$ H122α recombination line velocity of the H II region (Celnik 1985). Since the neutral and ionized gas share the same LSR velocity, this gives further support to the contention that this feature is an expanding shell.

The (ℓ, v) diagram also shows two local minima in the shell structure rather than a single minimum as one might expect. Also the shape of the expanding shell is far from being spherical. This is evident in Figure 1, and especially Figure 5, which is the intensity map integrated over the expansion velocities (11.3 to 20.6 $\mathrm{km\,s^{-1}}$) derived from the expanding shell model. The features in the maps are indications of the density inhomogeneities of the medium. Apparent, too, in Figure 5 are the two areas of low column density, located at $(206°465, -2°083)$ and $(206°065, -2°283)$. The relative minima in brightness temperature occur at velocities 16.5 and 15.5 $\mathrm{km\,s^{-1}}$, respectively. Thus Figures 1 and 5 show two possible expansion centers within a larger H I shell. Since the OB stars associated with NGC 2244 are spread over an area a few arcminutes across, one would not expect a single expansion center. Of the OB stars listed by Cox et al. (1990) only HD46223 $[(\alpha, \delta)_{1950} = (6^h29^m, +4°51'); (\ell, b) = (206°465, -2°045)]$ is near the center of one of the minima. The other OB stars are located between the minima. Without knowing the radial velocity of HD46223 it is not certain that the star is physically associated with the minimum.

3. Summary and Conclusions

A study of the atomic gas distribution in the direction of the Rosette complex has achieved two major results:

(1) a new, high resolution H I map showing the distribution of atomic gas in the Rosette Molecular Complex and

(2) the measurement of the expansion of the nebula into the surrounding medium.

The total H I mass mapped in an 8.4 square degree area is $1.5\pm0.1 \times 10^5 M_\odot$, for optically thin emission.

Can the central star cluster provide sufficient mechanical energy to power the observed expansion? Stars with spectral types earlier than B2 are known to have strong

stellar winds (Castor *et al.*, 1975) which are expected to impart momentum to the medium. Cox et al. (1990) list four O stars and nine other stars earlier than B2 as NGC 2244 cluster members. These spectral types are all expected to have significant winds throughout their lifetimes. The energy from Lyman continuum photons also can be transferred to the surrounding medium in the form of mechanical energy. Below we present order of magnitude calculations for the kinetic energy and momentum contained in the shell and compare these to the energy and momentum input expected from the NGC 2244 cluster stars.

To determine the magnitude of energy needed to accelerate the shell we need to estimate the total (atomic and molecular) mass of the shell. The molecular material surrounding the central cluster is clearly defined in CO maps such as that in Blitz & Thaddeus 1980. From the Bell Labs CO Survey of the area (Blitz & Stark 1986, Blitz 1991), we estimated the H_2 mass of the shell to be 5700 $M_\odot$ (see also Kuchar et al. 1993). The same boundaries defined by the CO emission were used to determine the H I mass of the shell. This amounts to 6600 $M_\odot$, with the total mass in the shell being approximately 1.23×10^4 $M_\odot$. The H I mass is probably underestimated since the H I shell may extend beyond the boundaries defined by the CO emission. From examining velocity-integrated CO maps of the region, the H_2 column density is approximately 20 to 30 % of the H I column density. Using the upper limit of 30% implies that the total H mass of the shell could be as much as 1.9×10^4 $M_\odot$. For the purpose of an order of magnitude calculation this mass will be used to determine the maximum momentum and energy required by the expansion.

The kinetic energy of the expansion is approximately 3.8×10^{48} ergs, where M is the total mass of the shell listed with an expansion velocity of ~ 4.5 km s^{-1}. Castor et al. (1975) estimate that the total mechanical energy available from stellar winds of O stars is of order 10^{50} ergs over the lifetime of the star. Since NGC 2244 contains four O stars, it can provide sufficient mechanical energy to power the expansion. We estimated the age of the cluster at approximately 4 Myr from dynamical considerations. However, others have estimated the cluster's age as young as 10^5 yrs (Raimond 1966). This would suggest that at most only 4% of the mechanical energy available from NGC 2244 went into accelerating the molecular gas.

Momentum transfer is the mechanism which imparts mechanical energy to an

expansion such as this. For a steady wind of velocity V_w and a stellar mass loss of $\dot{M}$ the rate of momentum injection into the surrounding medium is

$$\dot{M}V_w = M_s V_s / t_s, \tag{7}$$

where M_s, V_s, and t_s are, respectively, the mass, velocity, and dynamical timescale of the expanding shell. We estimated the dynamical timescale as 4×10^6 yrs. Thus the momentum transfer required for the shell is 2×10^{-2} $M_\odot$ yr^{-1} km s^{-1}. Weaver et al. (1977) estimated the momentum transfer from a typical O star as 2×10^{-3} $M_\odot$ yr^{-1} km s^{-1} (for $\dot{M} = 10^{-6} M_\odot$ yr^{-1} and $V_w = 2000$ km s^{-1}). From this calculation, 10 O stars would be needed to supply the necessary momentum. Although NGC 2244 has four O stars and several B stars, it is unlikely that the winds could supply *all* of the required momentum.

Since radiation from the Lyman continuum can also impart energy to the gas, we also need to estimate the magnitude of the energy available to accelerate the shell. For ionization limited H II regions, Lasker (1967) calculated the efficiency, ϵ, for transferring continuum energy to kinetic energy. Although he calculates ϵ for stars as early as O5, NGC 2244 contains an O4 star. The efficiencies tend to decrease for earlier spectral types, therefore an order of magnitude estimate would place ϵ between 0.01 and 0.001. All of the O and B stars can provide 1.23×10^6 $L_\odot$ beyond the Lyman continuum (Panagia 1973). If the age of the cluster is of order 1 Myr, the total energy available from the Lyman continuum over this period is 1.5×10^{53} ergs. Thus the radiation from the cluster can also provide more than sufficient energy to power the expansion.

The central cluster can provide sufficient energy to power the flow through a combination of stellar winds and radiation. The important consideration here is the efficiency of each process. Stellar winds and radiation contribute to some degree to the expansion process, since both are present in the complex. It is apparent from the order of magnitude calculations above that a very small percentage of the energy budget is needed to power the observed expansion of the nebula.

TABLE 1

Observational Parameters

Parameter	Value
ℓ–Coverage	
minimum	204°866
maximum	208°799
b–Coverage	
maximum	−1°016
minimum	−3°150
HPBW	4.′0
Sampling	4.′0
ΔV	0.52 $\mathrm{km\,s^{-1}\ channel^{-1}}$
ΔT_{rms}	0.4 K
Number of Spectra	1980

TABLE 2

Model Shell Parameters

Parameter	Value
R_s	48′
ΔR	4′
V_s	16 $\mathrm{km\,s^{-1}}$
V_{exp}	5 $\mathrm{km\,s^{-1}}$
T_s	100 K
τ_0	0.05
σ_v	2 $\mathrm{km\,s^{-1}}$

305

REFERENCES

Bania, T. M. and Kuchar T. A. 1993, in preparation

Bania, T. M. and Lockman, F. J. 1993, in preparation

Blitz, L. 1987, in Physical Processes in Interstellar Clouds, eds. Morfill and Scholer, (Reidel: Dordrecht), p. 35

Blitz, L. 1991, in The Physics of Star Formation and Early Stellar Evolution, eds C. J. Lada and N. D. Kylafis, (Kluwer: Dordrecht), p. 3

Blitz, L. and Stark, A. A. 1986, ApJL, 300, L89

Blitz, L. and Thaddeus, P. 1980, ApJ, 241, 676

Castor, J., McCray, R., Weaver, R. 1975, ApJL, 200, L107

Celnik, W. E. 1983, A&AS, 53, 403

Celnik, W. E. 1985, A&A, 144, 171

Celnik, W. E. 1986, A&A, 160, 287

Cox, J. P. and Guili R. T. 1968, Principles of Stellar Structure Vol. 2, (New York: Gordon and Breach). p. 944

Cox, P., Deharveng, L., and Leene, A. 1990, A&A, 230, 181

Goldsmith, P. F. 1988, in Molecular Clouds in the Milk Way and External Galaxies, eds. R. L. Dickman, R. L. Snell, and J. S. Young (Berlin: Springer-Verlag) p. 1

Kuchar, T. A., Bania, T. M., Blitz, L., and Stark, A. A. 1993 in preparaion

Lada, C. J. and Gautier III, T. N. 1982, ApJ, 261, 161

Lasker, B. M. 1967, ApJ, 149, 23

Mathews, W. G. and O'Dell C. R. 1969, ARA&A, 7, 67

Ogura, K. and Ishida, K., 1981, PASP, 33, 149

Panagia N. 1973, AJ, 78, 929

Perez, M. R., Thé , P. S., Westerlund, B. E. 1987, PASP, 99, 1050

Raimond, E. 1966, BAN, 18, 191

Reynolds R. J. and Ogden P. M. 1979, ApJ, 229, 942.

Weaver, R. McCray, R. Caster, J. Shapiro, P. and Moore, R. 1977, ApJ, 218, 377
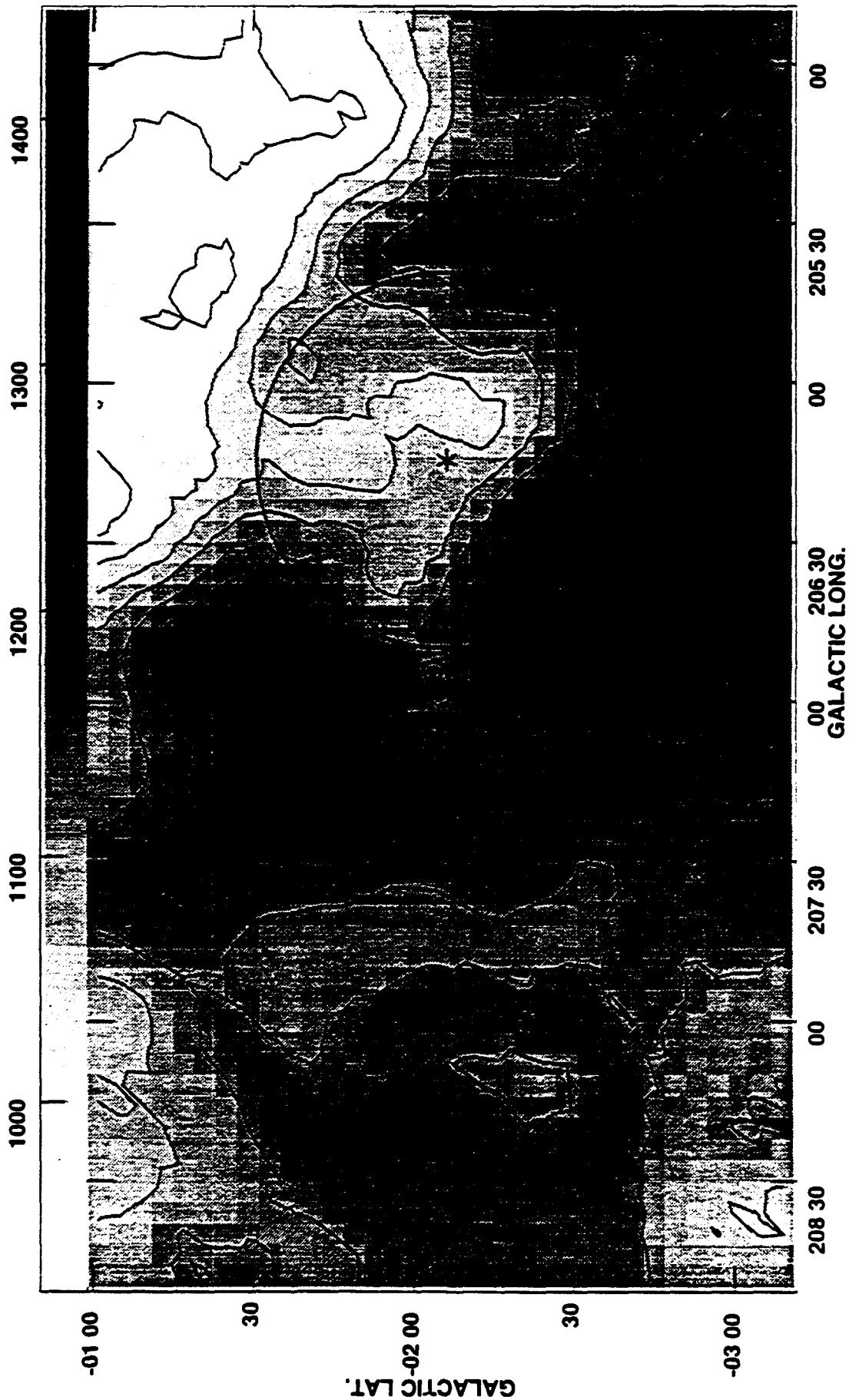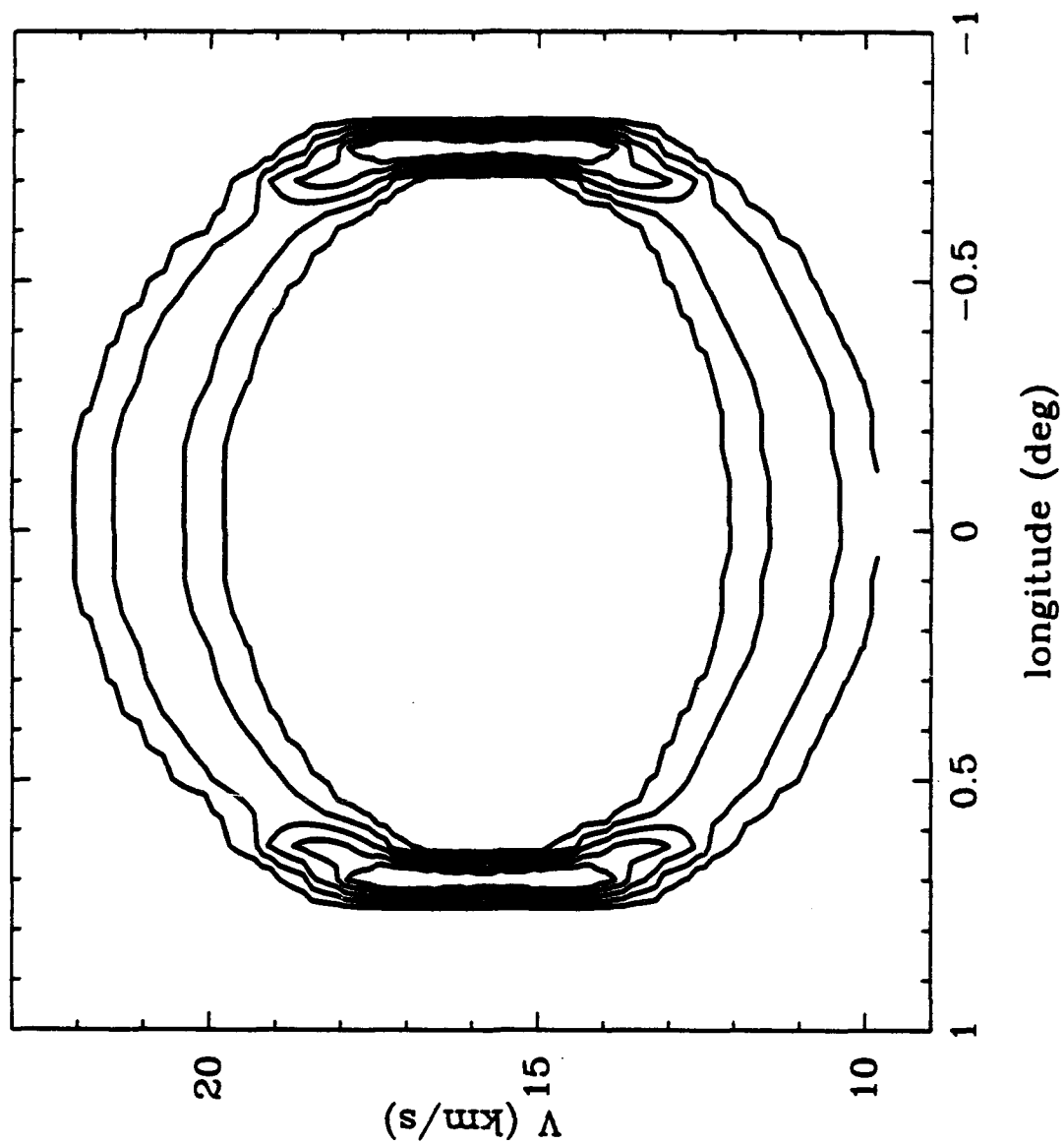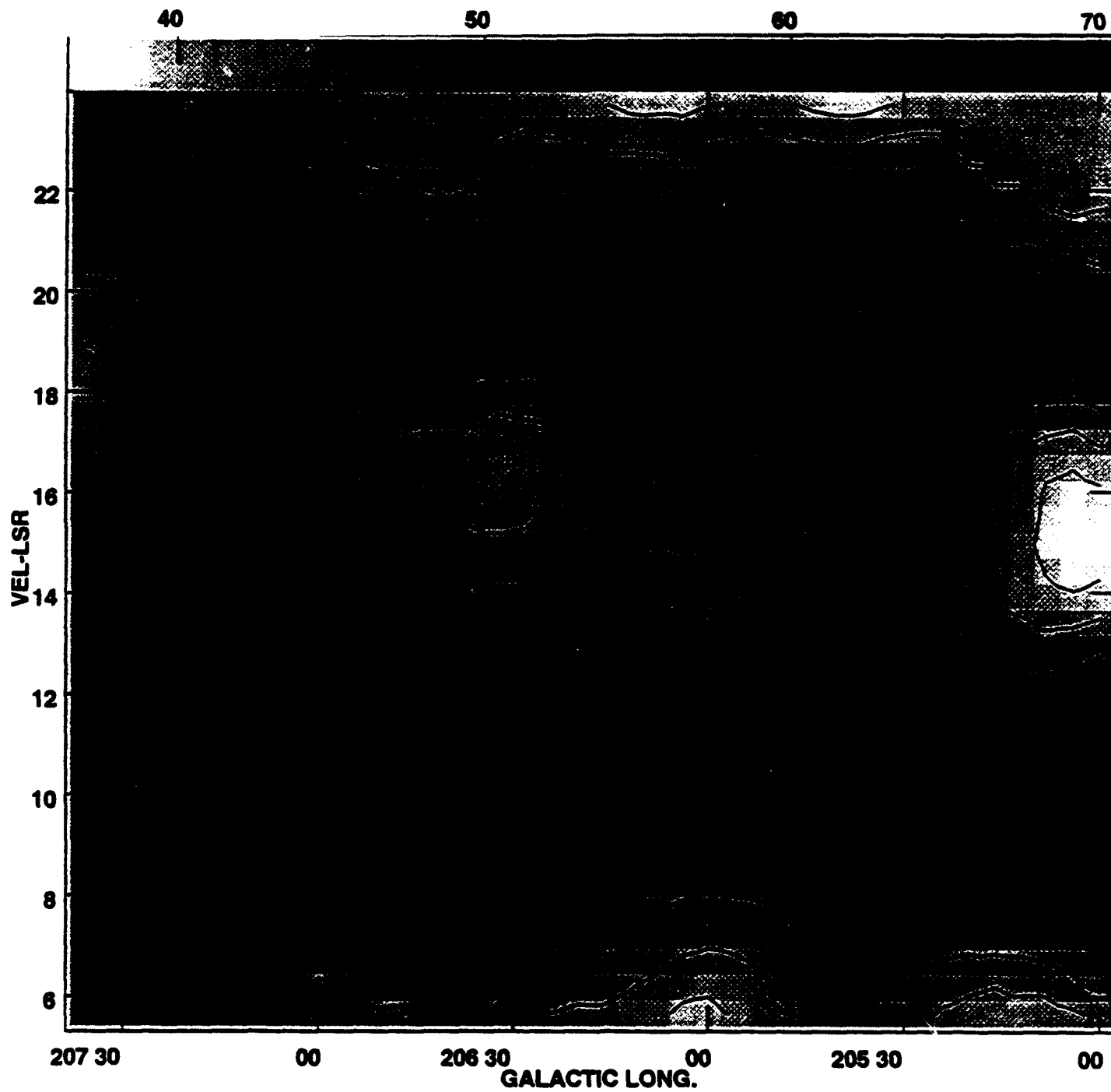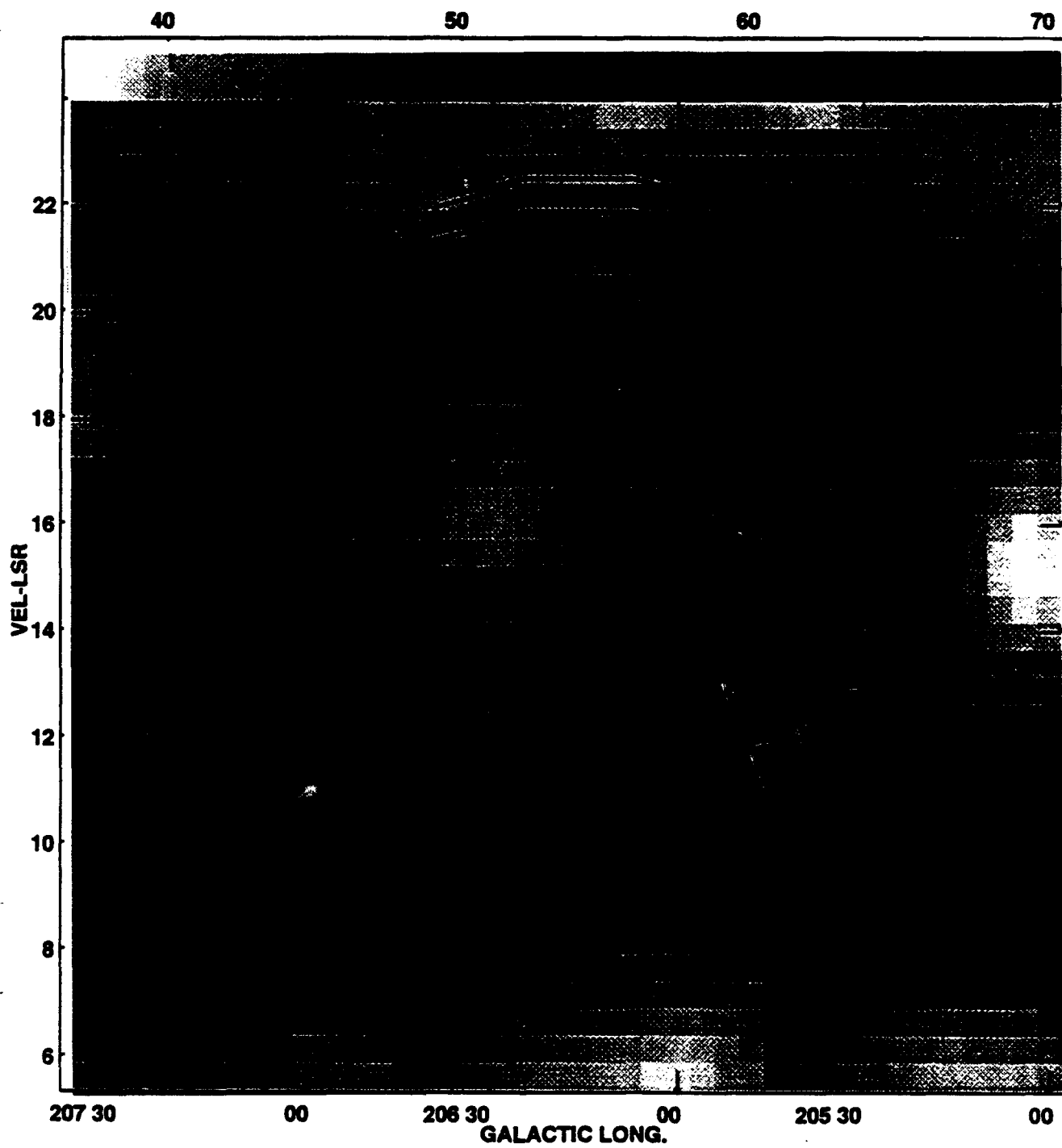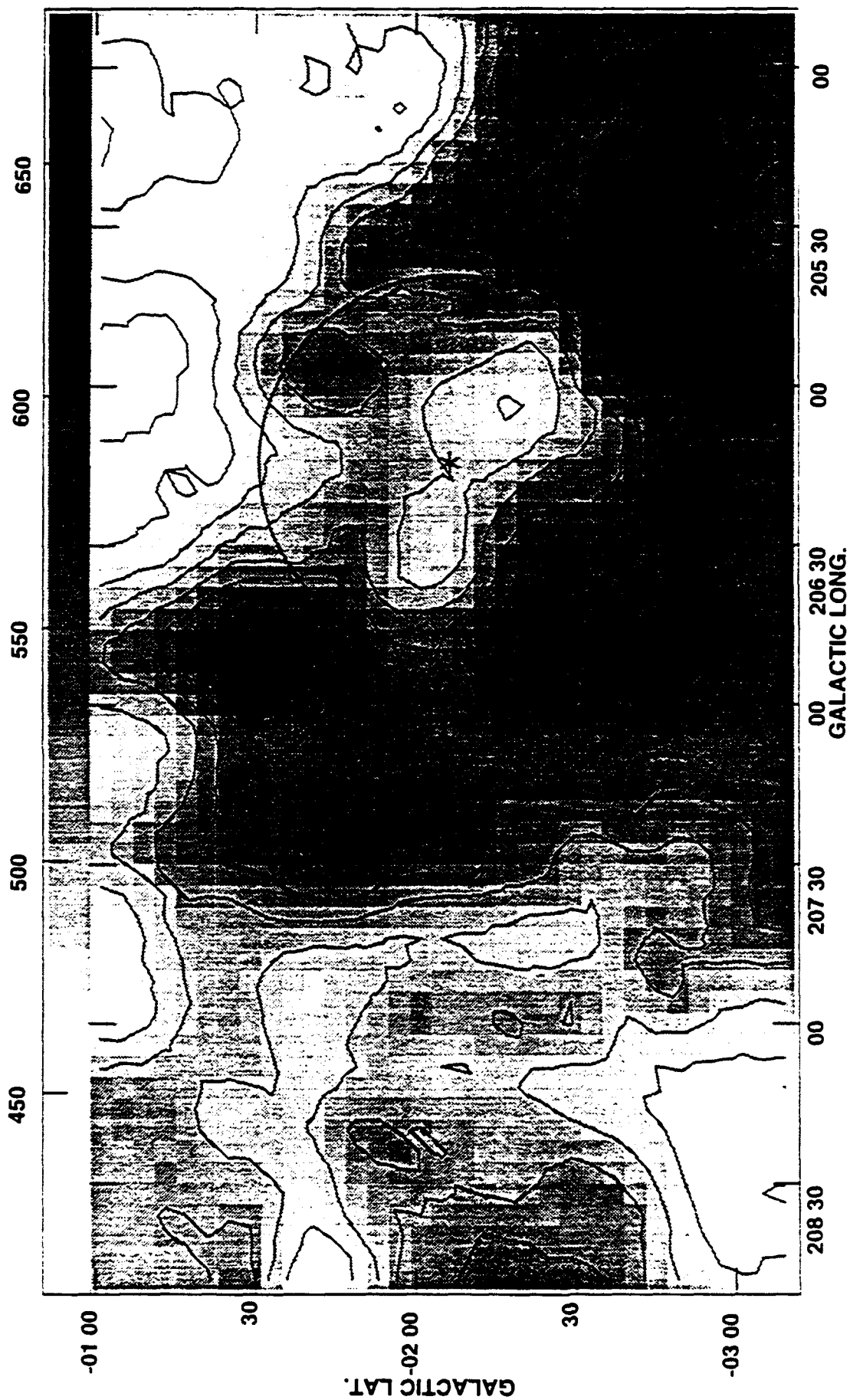
KINEMATIC DISTANCES OF GALACTIC H II REGIONS
FROM H I ABSORPTION STUDIES

THOMAS A. KUCHAR

Geophysics Directorate, Phillips Laboratory

29 Randolph Road, Hanscom AFB, MA 01731

and

T. M. BANIA

Department of Astronomy, Boston University

725 Commonwealth Avenue, Boston, MA 02215

Short Running Title: Distances of H II Regions

ABSTRACT

A catalog of kinematic distances for 70 Galactic H II regions is presented. H I absorption spectra were obtained toward 57 first quadrant ($30° < \ell < 60°$, $| b | \lesssim 0°5$) H II regions. Four of the sources have two recombination lines listed at their positions and thus the distances could not be fully determined. Only one spectrum had insufficient signal-to-noise to reliably detect H I absorption. Kinematic distances were determined for 51 H II regions. Combined with our previous study, the kinematic distances were determined for a total of 70 H II regions, 45 of which for the first time. The distribution of the H II regions peaks at a galactocentric distance of $R = 6$ kpc. Many of these sources are thus associated with star forming regions in the molecular ring.

Subject Headings:

Nebulae: H II Regions — Interstellar: matter — Radio Sources: 21 cm radiation

1. Introduction

As part of the Boston University – Arecibo Galactic H I Survey (Bania & Kuchar 1993), neutral hydrogen emission–absorption experiments were conducted toward a sample of 19 first quadrant H II regions confined to the Galactic plane (Kuchar & Bania 1990 hereafter referred to as Paper I). Paper I provided information about the spin temperature, optical depth, and column density of the absorbing gas along the line of sight to these H II regions. The absorption data also were used to set the kinematic distances to this sample.

The present paper extends this previous work by establishing the kinematic distances to 51 additional H II regions. The BU–Arecibo Survey was used as the database for the absorption study and was supplemented by pointed observations of the H II regions. The techniques used to detect reliable H I absorption are discussed in §2. These techniques were adapted from Paper I and Payne et al. (1980) and are only briefly discussed here. In §3 we present the observations for the absorption studies. The practical application of absorption studies to Galactic H II regions is reviewed in §4. Finally, the methods used to determine kinematic distances using H I absorption data are discussed in §5. The distances to all the H II regions observed for the BU–Arecibo Survey are also presented in this section.

2. Detecting H I Absorption

In these types of observations, cool H I is detected as the difference between a spectrum measured toward a source of continuum radiation, T_{on}, and the emission expected T_{off}, in that direction in the absence of a continuum source. Typically, T_{off} is an average of several spectra observed at positions nearby the continuum source. These positions are separated from the continuum source by the half power beamwidth (HPBW) of the antenna pattern. This separation is chosen in order to minimize the effect that the continuum source has when determining T_{off}, while still sampling emission from nearly the same line of sight.

Measuring both emission and absorption spectra toward a source of continuum radiation can reveal the physical properties of the absorbing gas. This was accomplished in Paper I. However, calibration problems in the intensity scale (see §3) required that the

analysis here be modified. As a result, the properties of the absorbing gas could not be reliably determined.

For frequency–switched spectra with baselines removed, the observables (T_{off} and T_{on}) are related to the optical depth, τ, of the absorbing gas and the continuum temperature, T_c, of the background source by:

$$\Delta T \equiv (T_{\text{off}} - T_{\text{on}}) = T_c(1 - e^{-\tau}). \qquad (1)$$

This equation assumes that the HI emission is from a thin slab of uniform properties (e.g. density and temperature). Also, these quantities are all implied functions of LSR (local standard of rest) velocity v.

The accuracy in detecting 21cm absorption is limited by the actual fluctuations of the HI brightness across the region of interest, since these fluctuations will produce uncertainty in the expected emission profile. Following the notation of Paper I and modifying the discussion for the present investigation, the RMS differences of the off–source spectra are defined as:

$$\sigma_{\Delta T}(v) = \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[T_i(v) - T_{i+1}(v) \right]^2 \right\}^{\frac{1}{2}}. \qquad (2a)$$

where N is the number of off–source spectra and the i subscripts denote the nearest neighbor off-source spectrum. Thus equation (2a) measures emission fluctuations over beam-sized angular scales. For a positive detection, it is required that $\Delta T > \sigma_{\Delta T}$.

Because receiver noise can also behave like emission fluctuations, a lower limit is placed on $\sigma_{\Delta T}$ at all velocities other than its value derived from equation (2a). Noise levels in the baselines of the spectra are used as estimates of these fluctuations. Any value of $\sigma_{\Delta T}$ derived from equation (2a) which is smaller than these baseline fluctuations cannot be considered a reliable estimate of the RMS fluctuations. Therefore a lower limit to $\sigma_{\Delta T}$ is based on measurements of the mean and standard deviation of both $\sigma_{\Delta T}$ and ΔT, evaluated in the baselines. Modifying the Paper I discussion here, we require that the lower limit of $\sigma_{\Delta T}$ be the *larger* of either (i) the sum of the mean value of $\sigma_{\Delta T}$ and twice its standard deviation evaluated in the baselines or (ii) five times the standard deviation of the ΔT, also evaluated in the baselines. The resulting reliability limit $\sigma'_{\Delta T}$ at each velocity channel is defined to be:

$$\sigma'_{\Delta T}(v) = \max\left[\sigma_{\Delta T}(v), \ <\sigma_{\Delta T}> +2\sigma(\sigma_{\Delta T}), \ 5\sigma(\Delta T) \right]. \qquad (2b)$$

This reliability limit is then a measure of fluctuations due to both spatial variations in H I emission and receiver noise.

3. Observations

The H II regions for this investigation were chosen from the Lockman (1989) recombination line survey. The sample was limited to 76 H II regions contained within the boundaries of the BU–Arecibo Survey (Bania & Kuchar 1993, hereafter referred to as the Survey): $30° < \ell < 60°$ and $| b | \lesssim 0°.5$. Pointed, high signal-to-noise spectra for 53 sources were obtained in 1990 July (Epoch 8 of the Survey) at the Arecibo Observatory[1].

---

[1] Arecibo Observatory is part of the National Astronomy and Ionosphere Center (NAIC) which is operated by Cornell University under contract with the National Science Foundation.

These pointed observations are used as the on–source spectra. The (ℓ, b, v) of H II regions observed within the Survey perimeter (including the 19 of Paper I) are given in Table 1. The other entries in the table will be discussed below. All spectra were observed with the 21cm 'flat' feed (HPBW = 4′) in frequency–switched mode with the reference frequency shifted 2.5 MHz from the H I line. The observations used the 1024 channel autocorrelator which provided a velocity resolution of 0.52 km s^{-1} per channel. However, these spectra could not be calibrated absolutely (see discussion below) and therefore were only useful for the determination of kinematic distances.

The Survey itself was used as the database for the off-source spectra. The first quadrant was sampled at 2′ for $\ell \lesssim 39°$ and at 4′ above that. Therefore the Survey can provide a sufficient number of off–source spectra for all of the 53 pointed observations. The Survey spectra have the same velocity resolution as do the pointed observations and individual spectra have typical RMS sensitivities of 0.5 K (in brightness temperature units). The criteria for choosing the off-source spectra are outlined in §4. Thus the number of

Survey spectra contributing to T_{off} ranged from a maximum of eight to as few as three. This produced RMS noise fluctuations (ΔT_{rms}) in T_{off} ranging from 0.30 to 0.17 K. Integration times for the pointed spectra ranged from five to 10 minutes providing ΔT_{rms} from 0.20 to 0.15 K (after being rescaled, as discussed below).

For the 1990 July observations, the local oscillator (LO) was randomly out of frequency lock because of an intermittent chip that controlled the frequency in the LO synthesizer. Such a problem could affect (i) the system temperature calculation, thus corrupting the intensity scale, and (ii) the frequency scale, hence corrupting the velocity scale.

The first problem caused the system temperature to vary dramatically ($\Delta T_{sys} \sim$ a few $\times$ 100 K). This resulted in large antenna temperatures that were as often negative as positive. To circumvent this problem, 10 one-minute spectra were taken toward each source. Spectra that were obviously affected by the LO fluctuations or had inconsistent system temperatures were discarded. (Inconsistent here refers to observations that did not follow the typical changes of T_{sys} with zenith angle.) The remaining spectra for an individual source had a linear baseline removed, were gain corrected, and then averaged together.

To determine if the LO problem caused any subtle variations in the calibration, the averaged spectra were compared to the Survey spectra corresponding to the position nearest the H II region. The velocity integrated intensities and system temperatures for the pointed and Survey data were compared. The system temperatures were corrected for variations in zenith angle by subtracting the Survey empirical system temperature curve from the observed system temperatures of both data sets. The result is an estimate of the continuum temperature, T_c. Since the survey data in general are not positioned directly toward the H II regions, their continuum temperatures are expected to be lower than those of the pointed observations. Also the velocity integrated intensity is expected to be larger for the Survey data, since they do not experience the same degree of absorption as the pointed data.

These comparisons showed that the pointed data had consistently lower continuum temperatures, by 20%, than the Survey data. However the integrated intensities for the pointed spectra were on average 2% larger than the Survey data. This is the opposite of what was expected. This same comparison was made between the Paper I observations

and the Survey data, since these data did not experience these LO fluctuations. The Paper I data comparisons were consistent with the expectations: the pointed observations had larger T_c and the integrated intensities were smaller by $\sim 5\%$.

Systematic comparisons were made between the pointed and Survey data to determine if the pointed data could be re-calibrated. The system temperatures and integrated intensities for both data sets were compared. These comparisons were made by taking into account the angular distance between the pointed spectra and the corresponding Survey spectra, since the farther a spectrum is observed from the source of continuum radiation the lower the value of T_c. However, no systematic trend was discovered in the data. It was concluded that any LO fluctuations that affected the pointed data were purely random. As a result, only a simple correction was made. The antenna temperatures of the pointed data were scaled to the ratio of velocity integrated intensities of the pointed data and the corresponding Survey data. This assured that the pointed data had at least the same column density (for optically thin emission) as the Survey data.

The accuracy of the velocity scale was checked by also comparing the pointed data with the nearest Survey observation. The two spectra were plotted together to see if there were any obvious velocity shifts. No such shifts were obvious to the eye.

A more quantitative analysis was done by taking the difference of these two spectra, averaging it over all velocity channels, and comparing how the mean and its RMS scatter vary when the velocity channels are shifted. The velocity scale of the pointed data was shifted by several velocity channels first and then subtracted from the Survey data. The average of this difference spectrum over all velocity channels is comparable to the difference of the velocity integrated intensities of both spectra and as such does not vary with the velocity channel shifts. The RMS of the mean, however, indicates how well the shapes, channel for channel, of the two spectra match.

The Paper I pointed data were compared to the Survey data as a check of this procedure. In each case, the RMS scatter was at a minimum for no velocity shift and increased when the pointed data were shifted to positive and negative velocities. This same comparison was made with the 1990 July data. With exception of four spectra, all of these data required no velocity shift. One spectrum, G59.8+0.2, showed a minimum in the RMS scatter over two velocity channels, corresponding to a zero and one channel shift. The four remaining spectra, G39.2−0.1, G41.2+0.4, G48.6+0.2, and G49.4−0.3,

showed minima in the RMS for a shift of just one velocity channel. These five spectra have velocity scales accurate to within in one velocity channel, whereas the majority appear to have been uncorrupted by the LO problem.

Without a more accurate method to calibrate the pointed observations, the derivation of the physical parameters of the absorbing gas were considered unreliable. However, the data are still useful inasmuch that they can be used to resolve the distance ambiguity for those H II regions (see §5), since frequency scale (and hence the absorption velocities) is accurate.

4. Analysis

There are several practical considerations in applying these techniques to Galactic H II regions. The foremost of these are: (*i*) the continuum emission is extended for H II regions rather than point-like; (*ii*) 21cm emission can emanate from behind the continuum source at Galactic LSR velocities which contribute to emission fluctuations; and (*iii*) the line of sight pathlengths traverse the Galactic plane and thus several H I clouds. Each of these must be taken into account not only when applying the experimental technique but also when interpreting the results.

A cursory examination of Galactic H II regions on a continuum map (e.g. Altenhoff et al. 1979, Reich et al. 1990) will reveal that many of the H II regions are extended sources which may be located in confused areas (i.e. nearby other continuum sources). The Survey mapping procedures take spectra at regularly spaced intervals without regard to the size of the continuum source. Therefore the spectra that are used to calculate T_{off} may overlap the continuum source and thus show some absorption. Since T_{off} is intended to represent the emission in the absence of a continuum source, care needs to be applied when choosing the spectra for T_{off}. For the H II regions that appear larger than the beamwidth of the telescope (4'), it is necessary to eliminate those spectra from T_{off}. Also, if the off-source spectra are observed nearby other strong continuum sources, then these too may have to be excluded. All of these spectra may show absorption against the continuum and can mimic emission fluctuations. This affects the value of $\sigma'_{\Delta T}(v)$ and the reliability of detected absorption.

Since these observations were made in the Galactic plane, there is necessarily 21 cm emission behind the source, some of which may be at the same LSR velocity as the H II region. This is not a problem if the continuum source is at or beyond the tangent point distance (see discussion in §5) or in the outer Galaxy. If this is not the case, background emission can complicate the interpretation of H I absorption in two ways. If the foreground H I clouds are cooler than those producing the background emission, then they can be seen in absorption against this emission (self-absorption) as well as against the H II region continuum. Thus some of the self-absorption features could be wrongly attributed to continuum absorption.

Secondly, background emission can contribute to the overall emission measured by T_{off}. Since the foreground clouds are the only clouds attenuating the continuum source, it is only their emission that is of interest. The H II regions in this sample are confined to $\sim 0°5$ of the plane in the first quadrant. Thus, the lines of sight practically cross the entire Galactic disk within the solar circle and, as a result, cross several H I clouds. If the background clouds are at the same Galactic velocities as the foreground clouds, then the background emission will contribute to an overestimate of T_{off}. Also, since T_{off} is an average of several spectra observed around the continuum source, some lines of sight will cross some clouds whereas other lines of sight will not. This tends to increase the measured emission, T_{off}, as well as the fluctuations, $\sigma'_{\Delta T}$. These fluctuations are typically about 5 K (Payne et al. 1982). The increased background fluctuations produced by these effects will complicate the analysis of 21cm absorption. Some artificial features might be produced while some real absorption might be masked. The criteria for detecting absorption was set to minimize the number of artificial detections. A consequence of this may be to miss some real absorbing clouds. In these cases the kinematic distance estimates to these H II regions may be unreliable.

5. Discussion: Kinematic Distances

Table 1 lists the 76 H II regions (57 pointed + 19 of Paper I), their Galactic coordinates and recombination line velocities from Lockman (1989, columns 2-4), H I absorption velocity associated with the H II region (column 5), the kinematic distances

based on the absorption data (column 6), and the resolution of the distance ambiguity (Near, Far, or Tangent point distances, column 7). The distances are based on the Clemens (1985) rotation curve. The kinematic distances of some of the HII regions were indeterminate based solely on HI absorption data. These sources are denoted by I in column (7) and are discussed below. Some of the HII regions are within the boundaries of the Westerhout (1958) catalog. These are denoted as such in column (8). This column also indicates if the distance ambiguity of the HII region had been previously resolved. These determinations were taken from a compilation by Wilson (1980), which contains a comprehensive list of first quadrant HII regions. There are some discrepancies between Wilson's list and the distances in column (6). These will be discussed below.

The two-fold distance ambiguity for HII regions can be resolved if HI is seen in absorption against the broad band thermal continuum of the HII region. The distance is then determined by observing the maximum velocity of the absorbing gas. If the HII region is at the near distance, then any HI gas beyond the HII region, some of which is likely to be at higher radial velocities, cannot absorb against the thermal continuum. Therefore an HI absorption spectrum will show absorption only up to the velocity of the HII region. If the HII region is located at or beyond the tangent point distance, then it is possible to see HI absorption beyond the velocity of the HII region and up to the tangent point velocity.

Figures 1, 2, and 3 show examples of HII regions identified at the tangent point and near and far kinematic distances, respectively. In each figure, T_{on}, T_{off}, ΔT and $\pm\sigma'_{\Delta T}$ are shown. Both the tangent point and recombination line velocities are flagged in each figure. For each case, note the relative positions of these two velocities and the maximum velocity at which HI absorption is reliably detected. These relative velocities are the basis for kinematic distances determination.

Two assumptions are implicit in this method of distance determination: (i) the HI is not seen in self–absorption and (i) there are no streaming motions associated with the rotation curve. For some of the sources placed at the far kinematic distance, the cool, foreground HI could be absorbing warmer background HI emission. In these cases, it is possible that the absorption seen beyond the HII region velocity may actually be HI self-absorption and not continuum absorption. The distance ambiguity thus remains.

The kinematic distances listed in Table 1 are again based up the Clemens (1985) rotation curve. (Paper I distances were based on the Burton & Gordon 1978 curve.) If

streaming and other non-circular motions are taken into account when determining the Galactic rotation curve, then the line of sight can cross several H I clouds that have the same radial velocity as the H II region (see e.g. Anantharamaiah et al. 1984). This complicates the procedure for determining distances using H I absorption (or any other absorbing tracer), since there are several distances along the line of sight that might be associated with these clouds and hence the H II region. However if the H I gas is not self-absorbed and circular rotation is assumed, then the kinematic distances are, on average, accurate to $\sim 20\%$.

In the analysis of the H I absorption features, we assume that neutral hydrogen is physically associated with the H II region, since dense H I clouds are often associated with H II regions (Greisen and Lockman 1978). The neutral hydrogen located just beyond the ionization front of the H II can absorb against the thermal continuum of the H II region. We expect to see H I absorption at or near the velocity of the H II region recombination line. There will likely be some velocity difference between the ionized and neutral gas due to the relative velocity of the ionization front with respect to the ambient gas. Therefore when considering H I absorption velocities associated with the H II region, we focus on a velocity range ($\sim \pm 10$ km s^{-1}) about the radio recombination line velocity.

Of the 76 H II regions in Table 1, 70 (92%) have detectable H I absorption within ~ 10 km s^{-1} of the H II region velocity. One source had no reliably detected absorption at all and the remainder had no reliably detected H I absorption within ~ 20 km s^{-1}. The H I absorption velocities (column 5 of Table 1) are compared to the recombination line velocities in Figure 4, in which a histogram of the velocity differences is shown. The histogram peaks at 0.6 km s^{-1}. For a random orientation of H II region streaming velocities with respect to the H I cloud, the distribution would be expected to peak at 0 km s^{-1}. However the peak is just slightly more than one velocity channel (0.52 km s^{-1}) away from zero velocity. Thus the difference is not too significant. Also, the dispersion of the histogram is 4.2 km s^{-1}. This value is close to the one dimensional velocity dispersion of 5 km s^{-1} which is often quoted for H I clouds (Crovisier 1978).

Five of the H II regions listed in Table 1 have no associated H I absorption within ± 20 km s^{-1} of the recombination line velocity. A typical example is illustrated in Figure 5. Since absorption was detected beyond this limit, it was still possible to assign kinematic distances to these H II regions. In these cases, it is possible that the H I associated with

these H II regions is behind the continuum source. Thus there would be no appreciable H I in front of the H II region nor H I absorption at or near the recombination line velocity. Each of these H II regions show a significantly detected 'negative' absorption feature, i.e. $\Delta T < 0$, at or near the recombination line velocity. Since this emission appears only in the on-source spectrum, it may well be the signature of H I lying *behind* the H II region.

Only one H II region, G34.9−0.0, did not show any absorption above the reliability limit. The uncertainty in the absorption was a result of the H II region being located near a strong continuum source. The Survey spectra which contributed to the calculation of T_{off} showed large emission fluctuations, which in turn contributed to a large $\sigma'_{\Delta T}$. Thus the kinematic distance to G34.9−0.0 remained indeterminate.

Several of the H II regions have recombination line velocities which exceeded the tangent point velocity for that longitude. The H I absorption features associated with these H II regions also have velocities which exceeded the tangent point velocity. The distances to these H II regions is very uncertain, since the non-circular motions may preclude them from having identifiable kinematic distances. By default, they were assigned to the tangent point distance. Although a few of these H II regions are at lower longitudes, most are in the 49° to 50° longitude range. The positions and velocities ($V \gtrsim 60\,km\,s^{-1}$) of these H II regions place them within the W51 complex. This is noted in column 8 of Table 1. Since these H II regions are within the same complex, they probably share the same distance as W51 even though they may not be physically at the tangent point distance.

Four of the H II regions in Table 1 (G30.5+0.0, G30.9+0.1, G31.1+0.1, and G30.2−0.1b) have two recombination lines observed at each of their positions. In each case, one of the recombination lines is at a low LSR velocity ($V < 50\,km\,s^{-1}$) while the second is at $V \simeq 100\,km\,s^{-1}$. The tangent point velocities for these longitudes are approximately $100\,km\,s^{-1}$, too. Thus the H II regions at the higher velocity which show H I absorption at $100\,km\,s^{-1}$(G30.9+0.1, G31.1+0.1, and G30.2−0.1b) were assigned to the tangent point distance. The kinematic distances for the lower velocity H II regions depend on whether these H II regions are responsible for absorption beyond their respective recombination line velocities. In all of the above cases, absorption is seen for $V > 50\,km\,s^{-1}$. However, it is unknown if this absorption is due solely to the H II region at the higher velocity or to both sources. The special circumstances here make the distances to the H II regions at the lower velocities indeterminate, since they can be at the near or far kinematic distances.

One of the distance discrepancies between Wilson's (1980) list and Table 1 involves the source G30.5+0.0. This position shows absorption at $92\,\mathrm{km\,s^{-1}}$, which we associate with the higher velocity H II region. However, there is absorption beyond this velocity at $106\,\mathrm{km\,s^{-1}}$, which is just beyond the tangent point velocity. This complicates the situation, since the absorption beyond $92\,\mathrm{km\,s^{-1}}$ indicates that one or both of the H II regions may be at the far kinematic distance. The distances for both entries at this position in Table 1 are thus listed as indeterminate. Wilson's list of H II regions is compiled in part from the Downes et al. (1980) H110α survey. Downes et al. list only one recombination line velocity for this position, the lower velocity ($V = 50.0\,\mathrm{km\,s^{-1}}$). As a result, Downes et al. have assigned the far kinematic distance to the H II region based on H_2CO absorption data. Given the above circumstances, the H I absorption data cannot support the far kinematic distance.

There are three other H II regions in Table 1 which have distances inconsistent with Wilson's list. These are G35.6−0.0, G49.4−0.3, and G49.4−0.2. The latter two sources are within the boundaries of the W51 complex. The H I absorption spectra for both H II regions clearly show absorption beyond the recombination line velocity, hence indicating the far kinematic distance. Of the eight sources in Table 1 that fall within the W51 boundary, these are the only two at the far kinematic distance. The remainder were placed at the tangent point. As was noted above, all eight of these H II regions may be at a common distance if they are physically associated with each other. Most of the W51 sources have velocities that depart from pure circular rotation and therefore were placed at the tangent point distance. However, G49.4−0.3 and G49.4−0.2 have velocities that are allowable for a purely circular rotation curve and thus were placed at the far kinematic distance. The entire complex may be at the far or tangent point distance, a difference in distance of ~ 2 kpc. Whether all of these H II regions are indeed at a common distance cannot be resolved here, since the distances listed in Table 1 are based *only* on H I absorption data and thus determined independently. Therefore sources which may be in the same complex may not have the same kinematic distance as listed in the table.

The final discrepancy involves G35.6−0.0, which Wilson lists at the near distance. The H I absorption data shows several features at velocities beyond the recombination line. These features extend up to the tangent point velocity, thereby placing this H II region at the far kinematic distance. Downes et al. determined that this H II region was at the

near kinematic distance, since no H_2CO absorption was detected beyond $V = 53\,\mathrm{km\,s^{-1}}$. However, the HI data contradicts this result, since absorption is clearly seen beyond the recombination line velocity.

Some of these contradictions with the Downes et al. distances can be understood if the clumpy nature of H_2CO is taken into consideration. Formaldehyde is collisionally excited into emission when densities in molecular clouds exceed 10^5 $\mathrm{cm^{-3}}$. If such densities do not exist along the entire line of sight, then distances based on H_2CO data can be misleading. The ubiquitous nature of HI allows the detection of absorption along the entire line of sight. Thus distances based on HI absorption data can be considered more reliable.

6. Summary

The positions of the HII regions relative to the Galactic center are shown in Figure 6. Here, a face-on view of the Galaxy is presented with fiducial symbols to mark the positions of the Sun, Galactic center, and Galactic longitudes. The distance of the Sun from the Galactic center is taken as 8.5 kpc. Although HII regions are expected to trace spiral structure, this is not quite apparent in the figure. The limited longitude coverage of the Survey may preclude the mapping of a complete spiral arm and so only arm segments may be visible. However, the distribution does seem to peak in a ring centered at a galactocentric radius of 6 kpc. This structure appears to mark the outer boundary of the molecular ring.

If the absorption observed here is solely due to cold gas attenuating the thermal continuum, these data can provide kinematic distance estimates for the HII regions. We established kinematic distances to 19 HII regions in Paper I and extended this list for 51 additional HII regions. Although a few of the distances to these HII regions remain ambiguous, kinematic distances were determined for a total of 70 HII regions, 45 of which were determined for the first time.

TABLE 1

KINEMATIC DISTANCES OF H II REGIONS

Source (1)	ℓ (deg) (2)	b (deg) (3)	V_{HII} (km s^{-1}) (4)	V_{abs} (km s^{-1}) (5)	d (kpc) (6)	(7)	Note[b] (8)
1 G30.5+0.0	30.539	0.024	46.1	49.5	—	I	F
2			95.2	91.2	—	I	W43
3 G30.6−0.1	30.602	−0.106	102.5	104.6	7.3	T	T,W43
4 G30.8−0.0	30.776	−0.029	91.6	94.3	5.7	N	N,W43
5 G30.8−0.2	30.832	−0.186	100.9	96.9	7.3	T	W43
6 G30.9+0.1	30.854	0.134	38.1	37.6	—	I	
7			100.0	102.5	7.3	T	W43
8 G31.0+0.1	30.950	0.078	102.0	95.3	6.6	N	W43
9 G31.0−0.0	30.999	−0.038	101.2	105.1	8.0	F	W43
10 G31.1+0.1	31.054	0.079	24.1	22.7	—	I	
11			99.1	103.1	7.3	T	W43
12 G31.2−0.1(a)	31.165	−0.127	41.4	41.2	12.0	F	
13 G31.2−0.1(b)	31.239	−0.108	29.9	39.2	—	I	
14			98.0	105.1	7.3	T	
15 G31.3+0.1	31.275	0.056	104.7	108.7	7.3	T	
16 G31.6+0.1	31.580	0.101	99.9	95.8	6.5	N	
17 G32.2+0.1	32.151	0.133	96.7	88.6	8.2	F	
18 G32.8+0.2	32.797	0.192	15.0	11.8	13.4	F	F
19 G33.1−0.1	33.129	−0.094	93.8	101.0	7.1	T	
20 G33.2−0.0	33.194	−0.010	100.9	95.3	7.1	T	
21 G33.4−0.0	33.418	−0.004	76.5	—	9.5	F	
22 G33.9+0.1[a]	33.914	0.111	101.4	106.7	7.0	T	T
23 G34.2+0.1	34.254	0.144	54.6	55.1	3.3	N	N
24 G34.9−0.0	34.932	−0.018	45.6	—	—	I	
25 G35.6+0.1	35.574	0.064	50.5	52.0	10.8	F	
26 G35.6−0.5[a]	35.588	−0.489	56.0	57.7	3.4	N	

[a] from Paper I

[b] distance from Wilson (1980)

TABLE 1

(*cont.*)

Source	ℓ	b	V_{HII}	V_{abs}	d		Note[b]
	(deg)	(deg)	(km s^{-1})	(km s^{-1})	(kpc)		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
27 G35.6−0.0	35.603	−0.033	52.8	52.0	10.7	F	N
28 G36.3+0.7[a]	36.289	0.734	76.5	79.4	4.8	N	
29 G36.5−0.2	36.459	−0.179	72.7	74.2	9.1	F	
30 G37.4−0.3[a]	37.361	−0.288	40.3	34.0	11.2	F	F
31 G37.4−0.1	37.370	−0.067	53.2	51.0	10.4	F	
32 G37.4−0.0	37.439	−0.040	53.0	53.6	10.4	F	F
33 G37.5−0.1[a]	37.538	−0.113	52.5	52.5	10.4	F	F
34 G37.6−0.1	37.636	−0.113	52.3	49.5	10.4	F	
35 G37.7+0.1	37.671	0.132	88.9	88.6	6.7	T	T
36 G37.7−0.1	37.749	−0.109	57.8	—	10.0	F	
37 G37.8−0.2	37.763	−0.216	65.3	63.5	9.4	F	F
38 G37.9−0.4[a]	37.871	−0.399	60.8	63.9	9.7	F	F
39 G38.1−0.0	38.051	−0.042	58.3	—	9.9	F	
40 G39.2−0.1	39.252	−0.056	22.6	33.0	11.9	F	
41 G41.1−0.2[a]	41.096	−0.213	59.4	62.8	9.1	F	
42 G41.2+0.4	41.235	0.367	71.3	69.6	4.8	N	
43 G41.5+0.0[a]	41.517	0.033	17.7	13.9	11.6	F	
44 G42.1−0.6[a]	42.108	−0.623	66.0	68.0	4.3	N	
45 G42.4−0.3[a]	42.431	−0.264	62.7	61.8	8.5	F	
46 G42.6−0.1[a]	42.568	−0.143	67.2	71.1	4.5	N	
47 G43.2+0.0	43.169	0.002	8.9	5.7	11.9	F	F,W49
48 G43.9−0.8[a]	43.890	−0.790	55.0	54.6	8.8	F	
49 G44.3+0.1	44.264	0.100	59.6	53.1	3.9	N	
50 G45.1+0.1	45.125	0.136	57.6	54.6	8.3	F	F
51 G45.4+0.1	45.451	0.060	55.9	56.2	8.3	F	F
52 G45.5+0.1	45.475	0.130	56.0	52.0	8.3	F	F

[a] from Paper I

[b] distance from Wilson (1980)

TABLE 1

(cont.).

Source (1)	ℓ (deg) (2)	b (deg) (3)	V_{HII} (km s^{-1}) (4)	V_{abs} (km s^{-1}) (5)	d (kpc) (6)	(7)	Note[b] (8)
53 G45.8−0.3[a]	45.824	−0.290	62.3	62.9	7.5	F	
54 G46.5− .2[a]	46.495	−0.247	57.2	56.6	3.8	N	
55 G48.6+0.0	48.596	0.042	17.2	18.0	10.3	F	F
56 G48.6+0.2	48.642	0.227	10.6	8.2	10.6	F	
57 G48.9−0.3	48.930	−0.286	66.5	58.7	5.6	T	T,W51
58 G49.0−0.3	48.997	−0.295	65.5	64.4	5.6	T	W51
59 G49.1−0.3	49.060	−0.260	64.1	60.8	5.6	T	T,W51
60 G49.1−0.4	49.076	−0.377	67.9	62.4	5.6	T	T,W51
61 G49.2−0.3[a]	49.204	−0.345	67.2	64.4	5.6	T	T,W51
62 G49.4−0.3	49.384	−0.298	53.7	51.0	7.3	F	T,W51
63 G49.4−0.2	49.407	−0.193	48.6	46.9	7.8	F	T,W51
64 G49.6−0.4	49.582	−0.381	62.1	58.2	5.5	T	T,W51
65 G50.0−0.1	50.024	−0.076	66.8	69.8	5.5	T	
66 G50.2+0.3[a]	50.232	0.326	71.2	63.4	5.4	T	
67 G51.1+0.2[a]	51.060	0.162	42.9	40.2	7.8	F	
68 G51.4−0.0	51.362	−0.001	52.7	56.2	4.0	N	
69 G52.2+0.7[a]	52.233	0.736	2.8	2.1	10.2	F	
70 G52.8+0.3[a]	52.753	0.335	16.6	—	9.3	F	
71 G53.2+0.2	53.184	0.155	8.3	7.2	9.7	F	F
72 G53.6+0.2	53.639	0.235	38.0	25.8	7.5	F	
73 G54.1−0.1	54.092	−0.066	42.1	46.9	6.9	F	
74 G57.5−0.3	57.541	−0.276	2.8	10.0	9.0	F	
75 G59.5−0.2	59.529	−0.181	29.4	31.3	6.3	F	
76 G59.8+0.2	59.796	0.237	−2.8	—	8.7	—	

[a] from Paper I

[b] distance from Wilson (1980)

REFERENCES

Anantharamaiah, K. R., Radhakrishnan, V., and Shaver P. A. 1984, A&A, 138, 131

Altenhoff, W. J., Downes, D., Pauls, T. A., Schraml, J. 1979, A&AS, 35, 23

Bania, T. M. and Kuchar, T. A. 1993, ApJS, in preparation

Burton W. B. and Gordon M. A. 1978, A&A, 63 7

Clemens, D. P. 1985, ApJ, 295, 422

Crovisier, J. 1978, A&A, 70, 43

Downes, D., Wilson, T. L., Bieging, J., and Wink, J. 1980 A&AS, 40, 379

Greisen, E. W. and Lockman, F. J. 1979, ApJ, 228, 740

Kuchar, T. A. and Bania, T. M. 1990 ApJ, 352, 192

Lockman, F. J. 1989, ApJS, 71, 469

Payne, H. E., Salpeter, E. E., and Terzian, Y. 1980, ApJ, 240, 499

——. 1982, ApJS, 48, 199

Reich, E., Reich, P., and Fürst, E. 1990, A&AS, 83, 539

Wilson, T. L. 1980 in *Radio Recombination Lines*, ed. P. A. Shaver (Reidel: Dordrecht) p. 205

Westerhout, G. 1958, Bull. Astr. Inst. Netherlands, 14, 215

FIGURE CAPTIONS

Figure 1: H I spectra showing an H II region identified at the tangent point distance. The top half of the figure shows the on (T_{on}, *dotted line*) and off (T_{off}, *solid line*) source spectra. The lower half shows the absorption spectrum (ΔT, *solid line*), along with its reliability estimate ($\pm \sigma'_{\Delta T}$, *dotted line*). The vertical lines flag the H II region velocity (*solid line*) and the tangent point velocity (*broken line*).

Figure 2: H I spectra showing an H II region identified at the near kinematic distance. See Figure 1 caption for a description.

Figure 3: H I spectra showing an H II region identified at the far kinematic distance. See Figure 1 caption for a description.

Figure 4: Distribution of velocity differences from Table 1. The differences between H II region recombination line velocities (column 4) and the H I absorption velocity (column 5) associated with the parent cloud of the H II region are displayed in histogram form. Velocity bins are $2\,\mathrm{km\,s^{-1}}$ wide.

Figure 5: H I spectra showing an H II region with no reliably detected absorption within $\sim \pm 20\,\mathrm{km\,s^{-1}}$ of the recombination line velocity. See Figure 1 caption for a description.

Figure 6: Distribution of H II regions in the Galaxy. This figure shows a face–on view of the Galaxy as seen from the north Galactic pole. The positions of the H II regions listed in Table 1 are plotted in galactocentric coordinates. The axes are labeled in kiloparsecs with fiducial symbols marks at the positions of the Sun ($\odot$) and Galactic center ($+$). Galactic longitudes are indicated as well. The locus of tangent points is shown as the broken, semicircular arc.

FIGURE 1

332
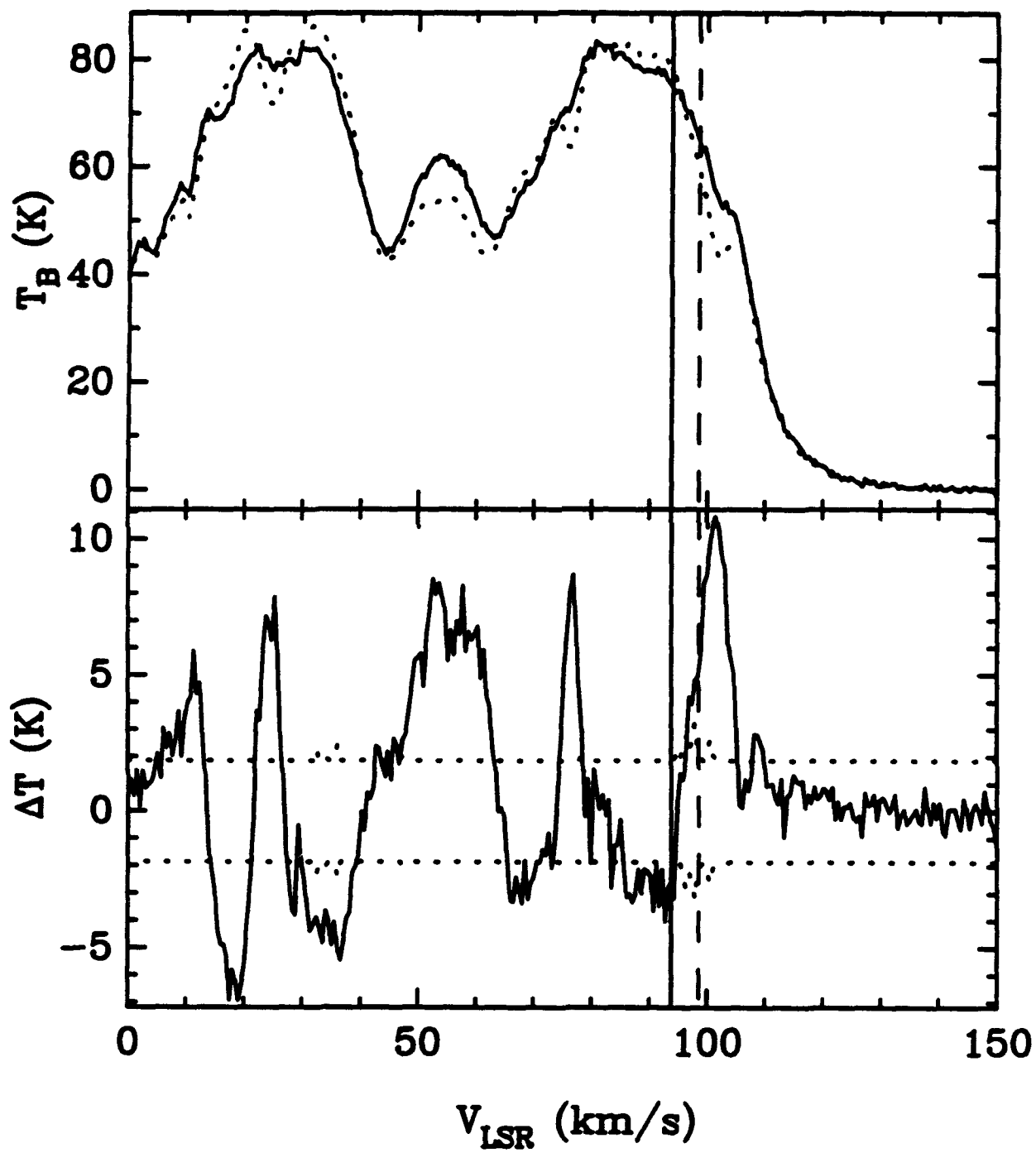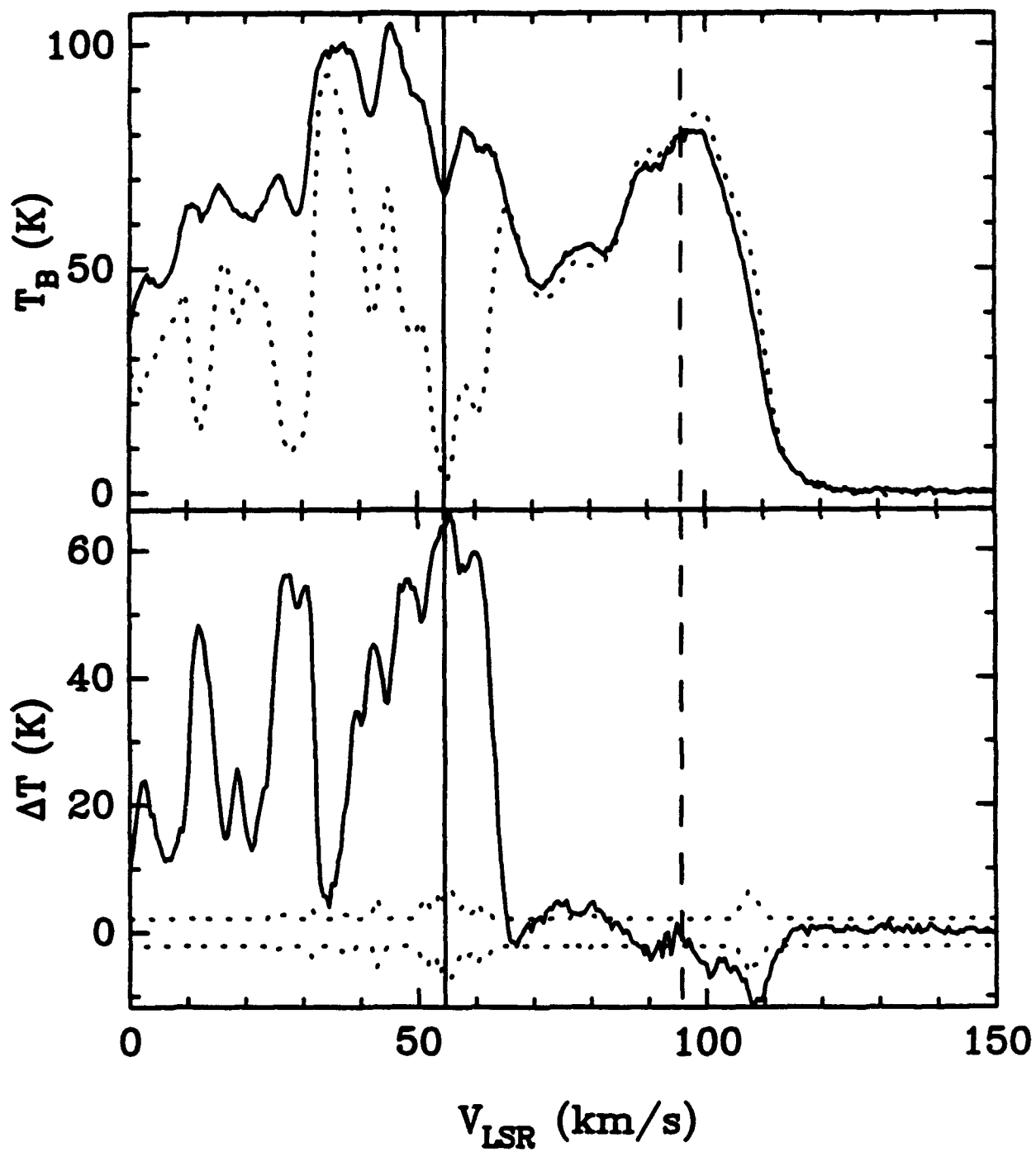
FIGURE 2

G37.8−0.2
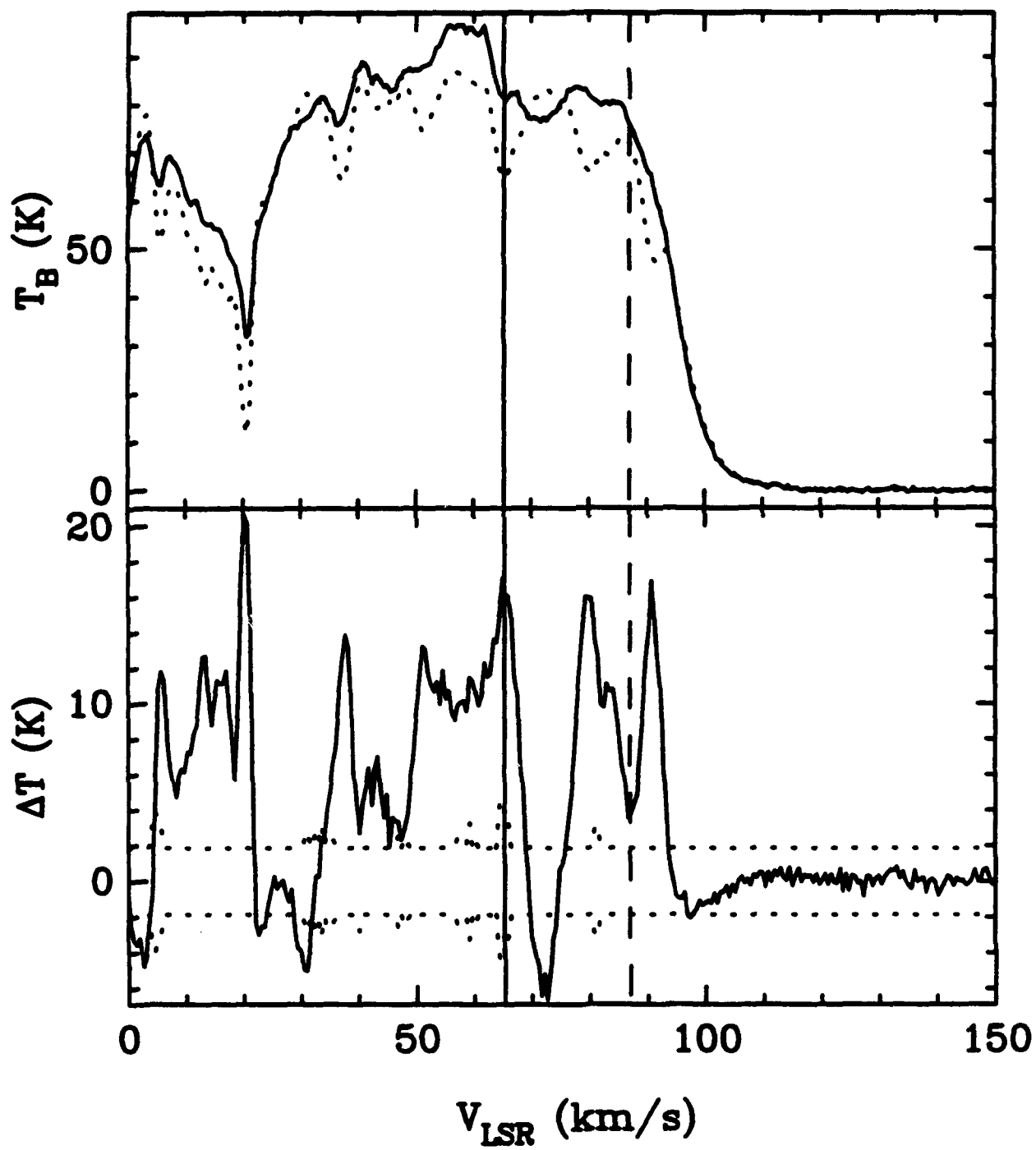
FIGURE 3
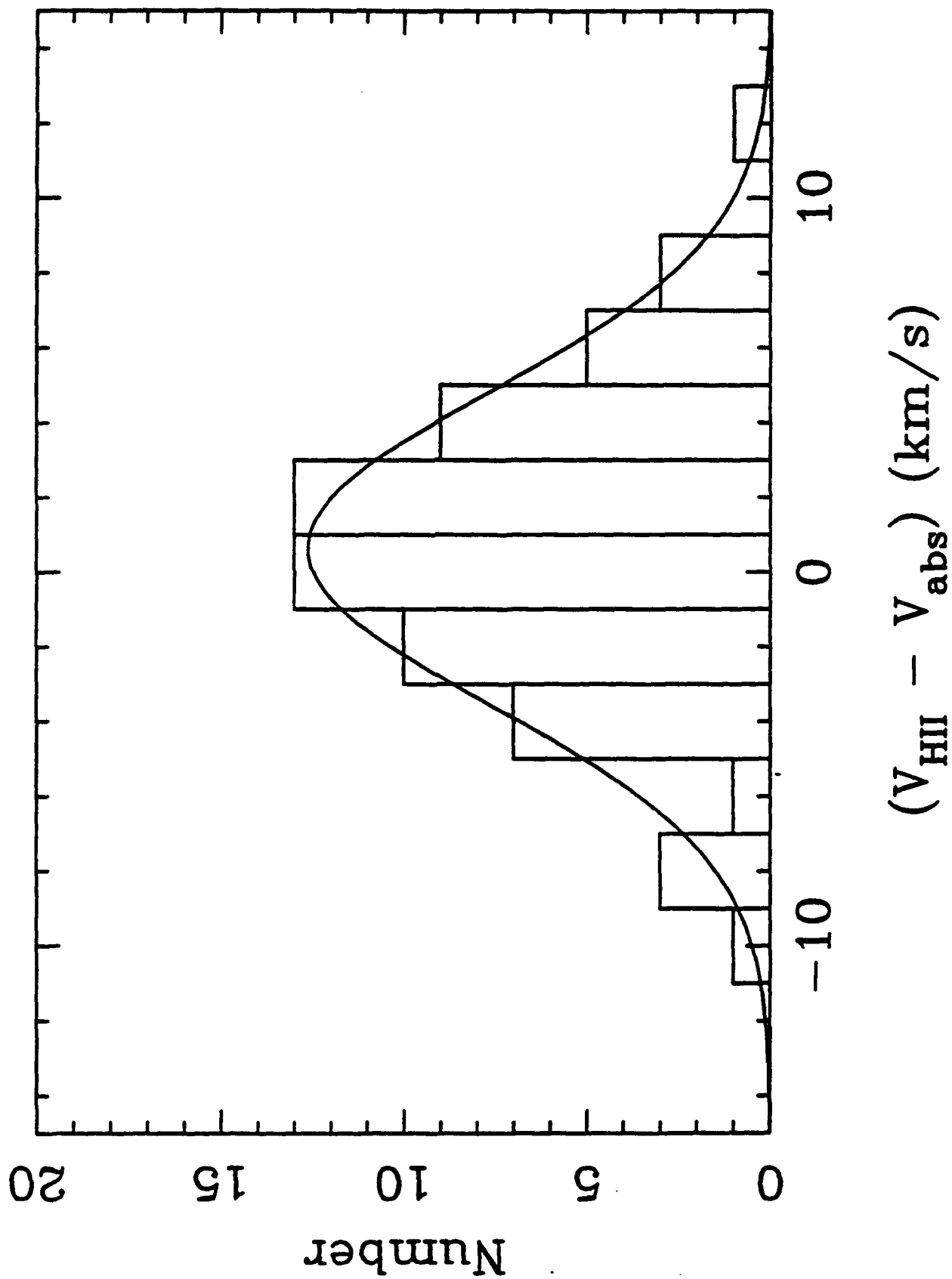
FIGURE 4

G38.1−0.0
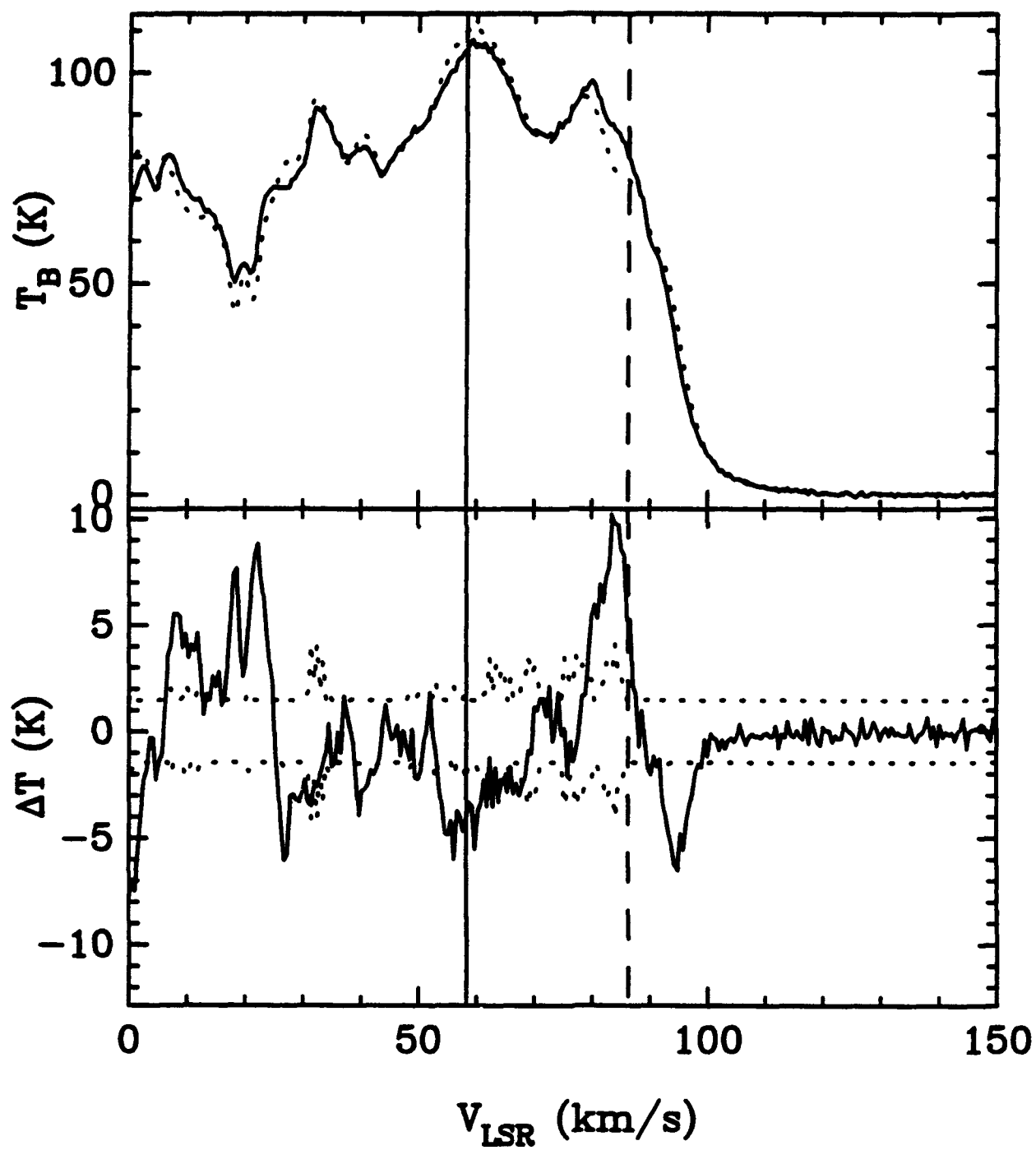
FIGURE 5
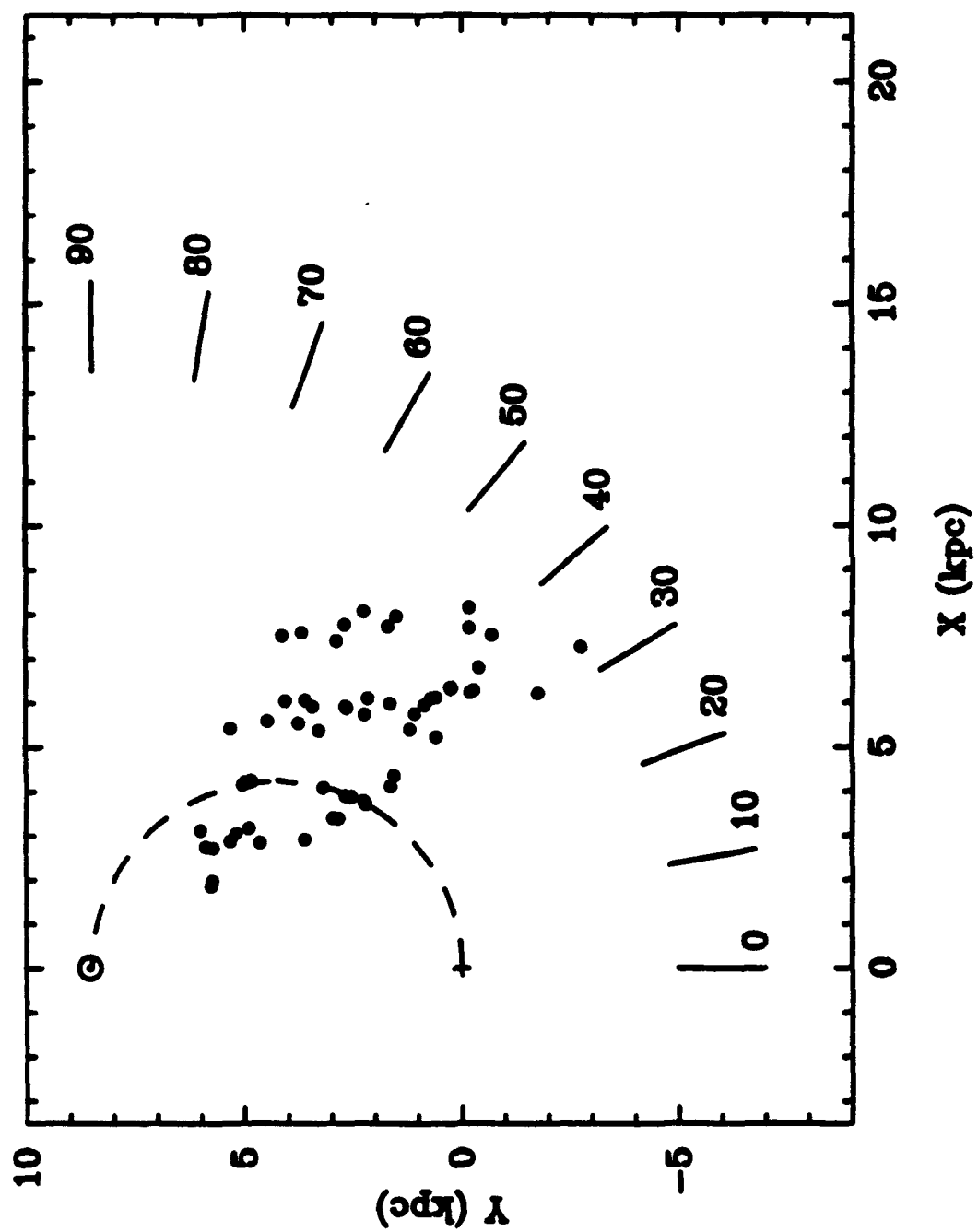
FIGURE 6

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

MODULATION OF THE HYDROXYL EMISSION BY A MONOCHROMATIC GRAVITY WAVE IN A REALISTIC NON-ISOTHERMAL ATMOSPHERE

Prepared by: Dr. Usama Makhlouf

Research Location: Phillips Laboratory,
                                Geophysics Directorate
                                Hanscom AFB, MA 01731

AF Research Colleague: Richard Picard

Contract Number: F19628-86-C-0224

Modulation of the hydroxyl emission by a monochromatic gravity wave in a realistic non-isothermal atmosphere.

by

Usama B. Makhlouf

Abstract

The effect of a realistic atmospheric temperature profile and realistic atmospheric photo-chemistry on gravity-wave -driven fluctuations in the OH nightglow is studied. The steady state OH vibrational distribution is calculated, which includes the effect of quenching by thermalizing collisions. A numerical model that includes a full wave treatment for waves propagating in an atmosphere with realistic temperature and background wind profiles is developed to generate the gravity waves that are used to modulate the [OH(v)], and hence the OH Meinel emission. The modulations in the OH airglow is quantified in terms of the Krassovsky ratio η, which will be compared with other theoretical models and some experimental results. Quenching is considered in this model and it has the effect of lowering η by about 20%, in addition to making η a vibrational dependent quantity. Inferring the temperature from airglow simulations was done in two ways to mimic experimental methods: a) using the ratio of two rotational lines in a vibrational level to infer the temperature (the rotational temperature) or b) using the Doppler width of a single line to infer the temperature (the Doppler temperature). Calculations by both these methods are compared with the commonly used brightness-weighted temperature. For gravity waves with small phase velocity (medium-scale waves) the differences in calculating η using these three methods are about 10%, while for gravity waves with large phase velocity (large scale waves) the η calculations using the ratio of two rotational lines agrees very well with the one calculated using the brightness weighted temperature, but the one from the Doppler width can be different by up to 50%.

Acknowledgment

1. Introduction:

Many observations of the hydroxyl airglow emissions, reveal evidence of temporal and spatial structures in these emissions [Krassovsky (1972), Peterson and Kieffaber (1973), Krassovsky et al (1977), Noxon (1978), Takahashi et al (1985), Taylor et al (1987), Viereck and Deehr (1989),Taylor and Hill (1991), Taylor et al (1991)]. These structures have been related to the passage of atmospheric gravity waves (AGW's) [Hines, (1960)] through the emission layer. The observations of these structures in the OH emission layer can be used as a probe to study the AGW's and their role in the dynamics of the middle atmosphere.

The Krassovsky ratio η, first introduced by Krassovsky (1972), relates the observed relative fluctuations in the airglow intensity to the relative fluctuations in the ambient temperature averaged over the emission layer. Nowadays η is widely used as a tool to quantify the effect of the AGW on the airglow. The η ratio has attained wide usage because it measures the response of photochemically generated airglow to a periodic dynamical perturbation such as an AGW in a manner independent of the amplitude of the perturbation, at least in the regime where the response is linear. In addition, it has been shown to reflect the airglow photochemistry and dynamics in a very useful manner. Krassovsky (1972) showed that η was determined simply by the order of the rate-limiting reaction and wrote:

$$\eta = \frac{2}{\gamma - 1} - \mu \qquad (1)$$

where

γ : specific heat ratio

μ : the temperature dependence $T^{-\mu}$, of the rate of the rate-limiting reaction, in this case the three body reaction for O_3 production.

When the AGW period is much longer than the photochemical lifetimes, the chemical production and loss rates are determined by the instantaneous species concentrations, and one would expect Krassovsky's expression for η, equation (1), to be approximately correct. However, when the AGW period shortens enough to become comparable to the photochemical lifetime, Walterscheid et al (1987) have shown that there is a very important wave-period-dependent modification of the simple Krassovsky result, due to coupling of the chemistry and dynamics.

When defining temperature for an inhomogeneous atmospheric emission layer, modelers have universally assumed that the brightness-weighted temperature (BWT) [Weinstock (1978)] is the quantity of interest. However, no instrument measures the BWT directly, and in this paper I compare η's based on the BWT to directly measurable temperature measures, namely the rotational temperature(ROT), based on the ratio of line intensities in a ro-vibrational band, and the Doppler temperature (DOP), based on the width of a single Doppler-broadened emission line. Finally, other researchers have tended to use generic unperturbed atmospheric profiles, including profiles for reactive species such as O_3, O, H, and HO_2, instead of profiles appropriate to the season, latitude, and time of day of the measurements to which they are comparing their model results. An exception is Hickey et al (1992), who have used minor-species profiles output from the photochemical-dynamical model of Garcia and Solomon (1985) for the month and latitude of interest. Since there can be a significant seasonal/latitudinal dependence to the unperturbed OH Meinel emission profile, as well as a diurnal variation, so in these calculations I choose profiles for the major-species in the background atmosphere from an existing climatological model [Summers and Anderson (1990)], then use a diurnal photochemical model to determine the profiles of the reactive minor species. I have adapted and updated a one-dimensional (1-D) diurnal photochemical model, to generate minor species profiles, Figure (1).

2. Vibrational-level production and loss processes:

The minor species profiles, generated from the 1-D diurnal model, are used as an input to the photochemical-dynamical model, which includes an expanded reaction set for production and loss of OH vibrational level populations [OH(v)]. This reaction set includes chemical quenching by atomic oxygen and collisional quenching by the major molecular species O_2 and N_2, in addition to multi-quantum radiative relaxation. The following reactions used in my photochemical-dynamical model;

(R1) $H + O_3 \xrightarrow{k_{1v}} OH(v) + O_2$ (v=6 - 9)

(R2) $O + HO_2 \xrightarrow{k_2} OH(v=0) + O_2$

(R3) $O + O_2 + M \xrightarrow{k_3} O_3 + M$

$$\text{(R4) } H + O_2 + M \xrightarrow{k_4} HO_2 + M$$

$$\text{(R5) } O_3 + HO_2 \xrightarrow{k_5} OH(v=0) + 2O_2$$

$$\text{(R6) } OH(v=0) + O_3 \xrightarrow{k_6} HO_2 + O_2$$

$$\text{(R7) } OH(v=0) + HO_2 \xrightarrow{k_7} H_2O + O_2$$

$$\text{(R8) } OH(v) + O \xrightarrow{k_{8v}} H + O_2 \qquad\qquad (v = 0 - 9)$$

$$\text{(R9) } O_2 + OH(v) \underset{k_{9v,R}}{\overset{k_{9v}}{\rightleftarrows}} OH(v-1) + O_2 \qquad\qquad (v = 1 - 9)$$

$$\text{(R10) } N_2 + OH(v) \underset{k_{10v,R}}{\overset{k_{10v}}{\rightleftarrows}} OH(v-1) + N_2 \qquad\qquad (v = 1 - 9)$$

$$\text{(R11) } OH(v) \xrightarrow{A_{v,v-n}} OH(v-n) + h\upsilon \qquad\qquad (v=1 - 9, n = 1 - 6, v-n \geq 0)$$
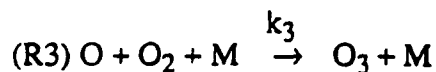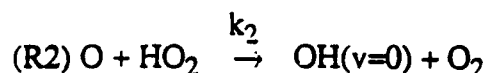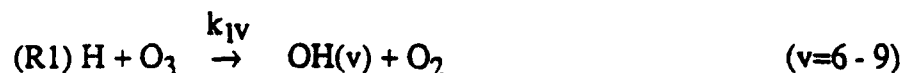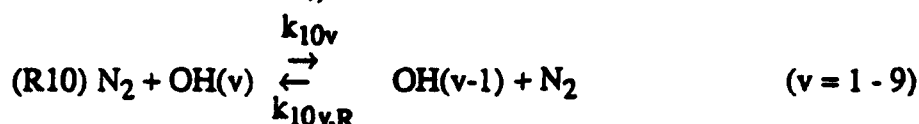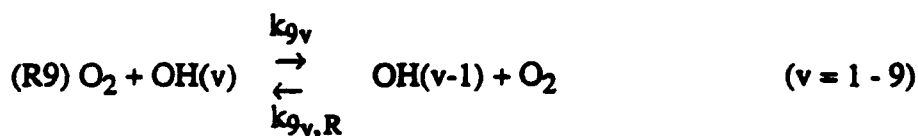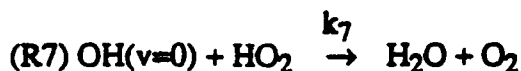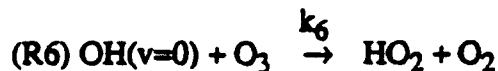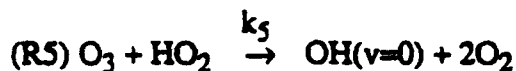
The rate constants $k_1,..., k_{10}, A_{v,v-n}$ are listed in Table 1.

The basic five reaction scheme of Winick (1983) for total [OH] employed by Walterscheid et al (1987) consists of reactions (R1) - (R4) and reaction (R8). I added quenching of OH(v) by O_2 and N_2 [reactions (R9) and (R10)], radiative decay [reaction (R11)] and additional sources and sinks of OH(v=0) [reactions (R5)-(R7)].

Since the work of Bates and Nicolet (1950), the chemical reaction (R1), between atomic hydrogen and ozone had been accepted as the primary reaction to produce excited OH(v). The vibrational levels v=6-9 were distributed according to Klenerman and Smith (1987) whose branching ratios were used. The secondary OH production reaction (R2) in my model is assumed to produce ground-level hydroxyl only [Lowe (1987)]. However significant differences in opinion exist in the literature concerning the role of this secondary reaction. For example Lopez-Moreno et al (1987) used this reaction to produce OH(v) in the v=3-6 vibrational states. On the other hand Kaye (1988) believes that this reaction produces OH in low vibrational levels v=0-3. Investigating all these possibilities, I found that whether this reaction is assumed to produce only ground state level

343

hydroxyl or lower or higher excited vibrational levels, it has no bearing on the calculated Krassovsky ratio.

Reactions (R5) - (R7) are added to the model in order to expand the reaction set and bring the sum of [OH(v)] calculated by the model into a closer agreement with the total [OH] calculated by the 1-D diurnal model.

The last reaction (R11) represents the radiative loss for the various vibrational levels, through single and multi-quantum transitions. The band averaged Einstein coefficients of Turnbull and Lowe (1989) were used in the calculations.

Many of the quenching rates required in the model are uncertain. Existing experimental measurements are incomplete, and there are few theoretical guides. Reaction (R9) represent the quenching of excited OH(v) by O_2, while reactions (R10) represent quenching by N_2. The quenching rates for molecular oxygen, reaction (R9), are taken from Dodd et al (1991) for v=1 - 6. Then, using a value from Sappey and Copeland (1990) for v=12, the remaining quenching rates for v=7 - 9 were obtained by logarithmically interpolating between v=1 and v=12. The quenching rates of the hydroxyl by nitrogen, reaction (R10), were obtained in a similar manner by interpolating and extrapolating between v=2 [Rensberger et al (1989)] and v=12 [Sappey and Copeland (1990)]. The rates of the reverse reactions, $k_{9v,R}$ and $k_{10v,R}$, resulting in collisional excitation, are determined from k_{9v} and k_{10v} by detailed balance; for example $k_{9v,R} = k_{9v} \exp[-(E_v-E_{v-1})/k_BT]$, where E_v is the energy of vibrational level v and k_B is Boltzman's constant. Reaction (R8) represents suddendeath chemical quenching by atomic oxygen. The rates for this reaction are somewhat controversial. For v=0 the rate was taken from J. Dodd et al (1990), and for v= 1 from Spencer and Glass (1977). Sivjee & Hamwey (19( ) chose 4.0×10^{-10} cm^3 sec^{-1} as the rate for v= 2 - 9 to fit their data. There are no experimental values available, and hesitating to use Sivjee and Hamwey's value since it exceeds the collisional limit (2.5×10^{-10} cm^3/sec) in the airglow layer region. A sensitivity study was done using a fast (2.5×10^{-10} cm^3/sec) and a slow (2.2×10^{-11} cm^3/sec) quenching rate for atomic oxygen for v=2 - 9. A large difference was found as a result of using the fast and slow rates, especially for the lower vibrational states. The results are illustrated in Figure (2), where it shows the effect of the choice of the rate constant k_8 on a high (v=9) and a low (v=3) energy level. Although atomic oxygen is not very abundant compared with N_2 and O_2 around the peak of the emission layer near 88 km, it is still an important quencher of OH, especially on the top side of the layer and for the lower vibrational levels, as seen in Figure (2). The effect on the lower levels is larger, since they are longer lived and, hence, more susceptible to quenching. Notice that the

[OH(v)] peak altitude is lowered as the quenching is increased. For my model calculations, I have chosen to use the fast rate $(2.5 \times 10^{-10}\,cm^3/sec)$ for $v = 2 - 9$, that is every collision of an OH(v) with atomic oxygen destroys the OH.

The Eulerian continuity equation for the [OH(v)] is given by:

$$\frac{\partial N_v}{\partial t} = Q_v - L_v \cdot N_v - div\,(N_v \vec{v}) \tag{2}$$

where

N_v : concentration of the species

Q_v : photochemical production rate

L_v : photochemical loss frequency

$\vec{v}$: gravity wave velocity field

The values of Q_v and L_v are determined from reactions (R1) - (R11), along with the rate constants in Table 1. These equations are coupled to similar continuity equations for the minor species involved in the OH photochemistry. In the remainder of this section, equation (2) is specialized to the steady-state, corresponding to no AGW modulation present, and examine the solutions to the resulting equations. I call this the zero order solution, since it is the solution without the AGW present.

In the steady-state, the divergence term in equation (2) vanishes, and the resulting equations describe photochemical equilibrium with production = loss:

$$0 = Q_{vo} - (L_{vo} \cdot N_{vo}) \tag{3}$$

whose solution is

$$[OH\,(v)\,] \equiv N_{vo} = \frac{Q_{vo}}{L_{vo}} \tag{4}$$

Using the reactions (R1) - (R11) and the results in Figure (1), I solve for N_{vo} (v=1-9) for the case mentioned above (day 269, 70°N, midnight).The results are shown in Figure(3). These results show (a) that the OH(v) vibrational distribution is not strongly altitude dependent, and (b)

345

that the peak of the vibrational levels v=1 - 9 fall within a 2 km altitude span with the v=9 peak is at a slightly higher altitude than the v=1, this is attributed to quenching by atomic oxygen especially on the top side of the layer, and since v=1 is longer lived than v=9, so the top side of v=1 is quenched down more effectively than v=9, see Figure (2).

From the level populations, Meinel-band volume emission rates ($I_{ovv'}$) can easily be calculated by multiplying by the corresponding Einstein coefficient, i. e.

$$I_{ovv'} = A_{vv'} \cdot [OH(v)] \tag{5}$$

In the next section, I will consider the AGW model and the linear response of the OH Meinel bands.

3. AGW Dynamical model and OH Meinel linear response:

Assuming that a monochromatic AGW perturbs the photochemical system described in section 2, and further assuming that the AGW amplitude is small enough so that the wave and the airglow response to it can be treated in the linear approximation, one can calculate the perturbation due to the AGW in the total molecular density N_{tot}, the temperature T and the horizontal and vertical components u and w of the velocity $\vec{V}$, using a realistic gravity wave model with background wind.

Starting with the linearized hydrodynamic equations:

$$i\Omega\Delta\rho + \Delta w \frac{\partial\rho_o}{\partial z} - ik_x\rho_o\Delta u + \rho_o\frac{\partial\Delta w}{\partial z} = 0 \quad\text{.......................(Continuity)} \tag{6}$$

$$i\Omega\rho_o\Delta u + \rho_o\frac{\partial V_{ox}}{\partial z}\Delta w = ik_x\Delta P \quad\text{...............(Horizontal momentum)} \tag{7}$$

$$i\Omega\rho_o\Delta w = -g\Delta\rho - \frac{\partial\Delta P}{\partial z} \quad\text{............................(Vertical momentum)} \tag{8}$$

$$\Omega\Delta P + \Delta w\frac{\partial P_o}{\partial z} = i\Omega c^2\Delta\rho + c^2\Delta w\frac{\partial\rho_o}{\partial z} \quad\text{..........................(Adiabatic)} \tag{9}$$

Combining equations (6) - (9) into two coupled equations to get:

$$\frac{\partial \Psi}{\partial z} - \eta \Psi = i\frac{\omega}{\Omega}(\omega_b^2 - \Omega^2)\phi \tag{10}$$

$$\frac{\partial \phi}{\partial z} + \left[\eta + \frac{k_x}{\Omega}\frac{\partial V_{ox}}{\partial z}\right]\phi = i\Omega\left[\frac{k_x^2}{\Omega^2} - \frac{1}{c^2}\right]\Psi \tag{11}$$

where

ω_b: Brunt frequency

V_{ox}: is the wind speed along the x-direction which is equal to

$V_{ox} = V_{we}\cos(\alpha) + V_{sn}\sin(\alpha)$

V_{we}: the zonal wind, V_{sn}: the meridional wind

α : the angle of propagation of the wave with respect to the East

Ω : Doppler shifted frequency

$$\eta = -\frac{1}{2}\left(\frac{g}{c^2} - \frac{\omega_b^2}{g}\right)$$

$$\Psi = \frac{\omega}{\sqrt{\rho_0}}\Delta P$$

$$\phi = \sqrt{\rho_0}\Delta w$$

Apply a radiation upper boundary condition at z_a=240 km, above which the atmosphere is assumed uniform, and so for a low altitude source: $\Psi \propto e^{ik_z z_a}$. Now, using Runge-Kutta-Fehlberg algorithm to numerically integrating the two coupled equations (10) and (11) from z_a to the ground, one gets a solution for $\Psi(z)$ and $\Phi(z)$. Using this solution and the linearized hydrodynamic equations (6) - (9), one can get the horizontal and vertical velocities of the gravity wave.

The next step is to linearize equation (2), along with similar continuity equations for the species O, O_3, H, and HO_2 about the zero-order steady-state solution of equation (4).Each vibrational level is considered as a separate chemically active specie so that I have a total of 14 species. The corresponding densities are denoted by N_i (i=0, 1,...., 13), corresponding to [OH(v)] (v=0-9), [O_3], [O], [H], and [HO_2]. The corresponding production and loss rates are denoted by Q_i and L_i. The number densities of O_2 and N_2 are assumed to be insensitive to photochemistry and affected

347

by the AGW dynamics only.

To first order, $T = T_0 + \Delta T$, $N_{tot} = N_{tot,0} + \Delta N_{tot}$, $N_i = N_{io} + \Delta N_i$, $Q_i = Q_{io} + \Delta Q_i$, and $L_i = L_{io} + \Delta L_i$, and all fluid parameters have space and time dependences $\exp[i(\Omega t - k_x x)]$. In the linear response model, all perturbed minor-species densities and related quantities will be proportional to this same factor. Assuming horizontal stratification, the first order linearized form equation corresponding to equation (2) is:

$$\frac{\partial \Delta N_i}{\partial t} = \Delta Q_i - N_{io}\Delta L_i - L_{io}\Delta N_i - N_{io}\left(\frac{\partial w}{\partial x} + \frac{\partial w}{\partial z}\right) - \frac{\partial N_{io}}{\partial z}w \qquad (12)$$

Letting all first-order quantities $\sim \exp[i(\omega t - k_x x)]$ and dividing by N_{io}, to obtain a set of coupled equations for the relative fluctuations in N_i:

$$(L_{io} + i\Omega)\frac{\Delta N_i}{N_0} = \frac{\Delta Q_i}{N_{io}} - \Delta L_i + ik_x u - \frac{\partial w}{\partial z} - \frac{1}{N_{io}}\frac{\partial N_{io}}{\partial z}w \qquad (13)$$

Substituting the zero-order number density, the perturbed production and loss rates for each species, and the AGW particle velocity in equation (13), to get 14 linear algebraic equations for the 14 unknowns of the perturbed species densities driven by the AGW velocity. Solving these equations to get $\Delta[OH(v)]$ (v=1-9), where $\Delta[OH(v)] = \Delta N_{i=v}$.

The perturbed volume emission rate $\Delta I_{vv'}$, is then defined as:

$$\Delta I_{vv'} = A_{vv'} \cdot \Delta[OH(v)] \qquad (14)$$

Dividing equation (14) by equation (5), to see that $\dfrac{\Delta I_{vv'}}{I_{vv'}} = \dfrac{\Delta N_v}{N_{vo}}$.

Since the A-coefficients have cancelled, then the fractional fluctuation in the emission rate is the same for all bands originating in the same upper level v. Using equation (13) for $\Delta N_v/N_{vo}$ and noting from equation (3) that $L_{vo}N_{vo} = Q_{vo}$, one can obtain an equation for the relative fluctuation in the emission rate;

$$\frac{\Delta I_{vv'}}{I_{vv'o}}\left(1 + \frac{i\Omega}{L_{vo}}\right) = \frac{\Delta Q_v}{Q_{vo}} - \frac{\Delta L_v}{L_{vo}} + \frac{1}{L_{vo}}\left(ik_x u - \frac{\partial w}{\partial z} - \frac{1}{OH(v)}\frac{\partial OH(v)}{\partial z}w\right) \qquad (15)$$

In the gravity-wave branch at least, $\omega < \omega_b$ =Brunt frequency $\approx 2\pi/[5 \text{ min.}]$, while L_{vo}

$\geq \sum_{v'} A_{vv'} \geq 22$ sec^{-1}, so that the second term on the left-hand side of equation (15) can be ignored in the equations of the hydroxyl vibrational levels [OH(v)]'s. This term can also be ignored in the equation for [HO$_2$] specie over most of the altitude range of interest. This is apparent from Figure (4) which is a plot of the chemical lifetime of the reactive species versus altitude. However, these terms will be retained since they do not add any complexities in solving these equations.

Equation (15) shows that the fractional first-order perturbed volume emission rate is a sum of three terms, which represent (i) production rate fluctuations $\Delta Q_v/Q_{vo}$, (ii) loss rate fluctuations $\Delta L_v/L_{vo}$, and (iii) fluctuations arising from the divergence term of the continuity equation, which represent advection and compression of [OH(v)]$_0$ by the AGW.

4. Inferring a temperature from the airglow emission:

The Krassovsky ratio η has been computed from airglow data and calculated from models in an inhomogeneous atmosphere, where the temperature associated with emission from a broad layer is not well defined.

We calculate η using temperature calculated by modelers and inferred by experimentalists using three different methods:

a) Theoretical modelers have typically used the brightness weighted temperature to calculate η. The brightness $B_{vv'}$, observed by a column-integrated airglow instrument is defined as the integral of the volume emission rate over the thickness of the layer:

$$B_{vv'} = \int I_{vv'}(z)\, dz = B_{ovv'} + \Delta B_{vv'} \tag{16}$$

where

$$I_{vv'} = I_{ovv'} + \Delta I_{vv'}$$
$$B_{ovv'} = \int I_{ovv'}(z)\, dz$$
$$\Delta B_{vv'} = \int \Delta I_{vv'}(z)\, dz$$

The fluctuation in the brightness weighted temperature is defined as:

$$\langle \Delta T \rangle_v = T_{1v} + T_{2v} - T_{3v} \tag{17}$$

where

$$T_{1v} = \frac{\int I_{ovv'}(z) \, \Delta T(z) \, dz}{B_{ovv'}}$$

$$T_{2v} = \frac{\int \Delta I_{vv'}(z) \, T_o(z) \, dz}{B_{ovv'}}$$

$$T_{3v} = T_{ov} \cdot \frac{\int \Delta I_{ovv'}(z) \, dz}{B_{ovv'}}$$

and

$$T_{ov} = \frac{\int I_{ovv'}(z) \, T_o(z) \, dz}{B_{ovv'}}$$

b) The rotational temperature, which is inferred by looking at a minimum of two rotational lines in one of the hydroxyl vibrational bands and assuming that these rotational lines conform to a Boltzman distribution, an average rotational temperature can be inferred.

In a realistic atmosphere, the ratio in the intensity of two rotational lines in a single OH vibrational band can be written as:

$$\frac{B_a}{B_b} = \frac{A_a g_a}{A_b g_b} \frac{\int\limits_z \left(\frac{N(z) \exp\left(-\frac{E_a}{k_B T(z)}\right)}{Z[T(z)]} \right) dz}{\int\limits_z \left(\frac{N(z) \exp\left(-\frac{E_b}{k_B T(z)}\right)}{Z[T(z)]} \right) dz} \tag{18}$$

where

A_a : is the ro-vibrational Einstein coefficient.

g_i : is the degeneracy factor of the specified rotational line.

k_B : is the Boltzman constant.

Z : is the partition function.

Using the assumption of a Boltzman distribution, one can now define an average inferred rotational temperature (T_{rot}) for the emitting region as:

$$T_{rot} = \frac{E_b - E_a}{k_B \ln\left(\frac{B_a A_b g_b}{B_b A_a g_a}\right)}$$
(19)

where B_a/B_b is calculated using equation (18), which is exactly the quantity measured by a ground based instrument tuned to the two rotational lines.

Assuming perturbation due to AGW, one can get the relative temperature fluctuation as a function of the difference in the relative fluctuation in the intensity of the two rotational lines being measured:

$$\frac{\Delta T_{rot}}{T_{rot}} = \frac{k_B T_{rot}}{(E_a - E_b)} \left\{ \frac{\Delta B_a}{B_a} - \frac{\Delta B_b}{B_b} \right\}$$
(20)

where $\Delta B_a/B_a$ is obtained by perturbing B_a in equation (18) and is given in Appendix I. Similarly $\Delta B_b/B_b$ is obtained.

Using equation (20) the relative temperature fluctuation is calculated, which in turn is used with the relative brightness fluctuation of the corresponding vibrational band from equation (16), to calculate the Krassovsky ratio η.

c) The Doppler temperature, which is inferred from the width at half the maximum of the intensity of a single rotational line in one of the hydroxyl vibrational band. The intensity of a single rotational line emitting from an altitude z_k with the assumption of a Gaussian distribution is given by:

$$I_0(\upsilon) = \frac{N_J(z) A}{\sqrt{\pi}\alpha_D} \exp\left[-\left(\frac{\upsilon - \upsilon_0}{\alpha_D}\right)^2 \right]$$
(21)

where

N_J : is the rotational population.

υ_0 : is the center frequency of the line.

α_D : is the line width given by:

$$\alpha_D = \frac{\upsilon_0}{c}\sqrt{\frac{2k_B T}{m}}$$
(22)

where

c   : speed of sound.

m   : molecular mass.

Using equation (21) and applying perturbation due to AGW one gets:

$$\frac{\Delta I(\upsilon)}{I_o(\upsilon)} = \left[\frac{(\upsilon - \upsilon_o)^2}{(\alpha_D)^2} - \frac{1}{2} \right] \frac{\Delta T(z)}{T(z)} + \frac{\Delta N_J(z)}{N_J(z)} \tag{23}$$

where $\dfrac{\Delta N_J(z)}{N_J(z)}$ is given in Appendix I.

For a ground based observation, the rotational line observed will be:

$$B(\upsilon) = \int (I_o(\upsilon) + \Delta I(\upsilon)) \, dz \tag{24}$$

and from the width of B(υ), the observed Doppler temperature is obtained. If I integrates equation (21) over the emitting layer, I will get the background line shape $B_o(\upsilon)$, and from the width, the background ambient temperature T_{Do} is inferred using equation (22).

From equation (24), which is the observed line, and using equation (22) again the observed temperature is inferred. Using the observed temperature and the background temperature the fluctuation in the temperature ΔT_D due to the AGW's is obtained.

To get the background brightness B_{Do} and the corresponding fluctuations in the brightness ΔB_D, I have to integrate over the frequency domain:

$$B_{Do} = \int B_o(\upsilon) \, d\upsilon \tag{25}$$

$$\Delta B_D = \left[\int B(\upsilon) \, d\upsilon \right] - B_{Do} \tag{26}$$

Finally the Krassovsky ratio η, is calculated using the three previous techniques for inferring the relative fluctuation in temperature:

a) Using brightness weighted temperature:

$$\eta_v = \frac{\Delta B_{vv'}/B_{ovv'}}{\langle \Delta T \rangle_v / T_{ov}} \tag{27}$$

b) Using rotational temperature:

$$\eta_{rot} = \frac{\Delta B_{vv'}/B_{ovv'}}{\Delta T_{rot}/T_{rot}} \tag{28}$$

c) Using Doppler temperature measurement:

$$\eta_{Dop}(Line) = \frac{\Delta B_D / B_{Do}}{\Delta T_D / T_{Do}} \tag{29}$$

Notice that in this case the relative brightness fluctuation in the line intensity is used to calculate the Krassovsky ratio. also I can calculate the Krassovsky ratio using the relative vibrational band intensity as follows:

$$\eta_{Dop}(Band) = \frac{\Delta B_{vv'}/B_{ovv'}}{\Delta T_D / T_{Do}} \tag{30}$$

The difference between calculating the Krassovsky ratio using equation (29) or (30) is very small in both amplitude and phase. So the major difference in the three previous methods in calculating the Krassovsky ratio is the way one infers the temperature.

Figure (5) shows the amplitude of η as a function of period, for a wave with a horizontal phase velocity of 70 m/sec, calculated using the brightness weighted temperature (BWT), rotational temperature (ROT) and the Doppler temperature (DOP). Figure (6) is the same as Figure (5), except a large horizontal phase velocity of 260 m/sec is used. Both Figures shows that η_{DOP} has the lowest values.

From equations (27) and (28), η_v and η_{rot} are complex quantities and the phase is inferred directly from the equation. Equation (29) on the other hand is real, and the phase of η_{Dop} is inferred by calculating the time evolution of the fluctuations in both the intensity and the temperature. Figure (7) shows plots of the phases for the (BWT), (ROT) and (DOP) cases, using a phase velocity of 70 m/sec. Figure (8) is the same as Figure (7), except a larger phase velocity of 260m/sec is used.

5. Effect of Quenching:

The effect of quenching on η is considerable, and quenching should be taken into consid-

353

eration especially when dealing with short period AGW's. Figures (9) shows the quenching effect on $|\eta(\tau)|$ calculated using (BWT), where τ is the wave period, for a low-lying (v=3), mid-level (v=6) and a highly excited (v=9) vibrational level for a horizontal phase speed, $v_{phx}=140$ m/sec. Without quenching $|\eta|$ is independent of v and when quenching is turned on $|\eta|$ is reduced by up to 20%, dependant on the period and on the vibrational band.

Quenching not only has effect on η, but it is important in determining the correct vibrational population profiles, since these profiles are treated as independent species in calculating the brightness fluctuation in the selected band. Without quenching the three solid lines in Figure (9), which represent the Krassovsky ratio of OH(3), OH(6) and OH(9) fall on the top of each other that is quenching adds a new dimension to the Krassovsky ratio, were it becomes vibrationally dependent.

6. Conclusion:

A photochemical-dynamical model for studying the OH airglow modulations due to atmospheric gravity waves have been developed. The model includes a realistic background temperature and a realistic background winds. Both the realistic temperature and background wind profiles have large effects on airglow response parametrized by the Krassovsky ratio. Figure (10) shows the effect of the wind on η, for waves with a 2 hour periods. Quenching has about 20% effect on lowering η, in addition to making it vibrational dependent quantity, and that is important since experimentalist tune their measurements to specific rotational lines in a specific vibrational band, and that raises the other important part of trying to infer the temperature in a way that mimic experimental methods, which had been shown to have a substantial effect on η.

Finally, in order to compare OH airglow AGW's response data to models, it is imperative to have a simultaneous measurements of the local background wind profile in addition to the AGW's parameters. The direction of propagation of the gravity wave is an important parameter that needs to be measured during observations in order to compute the correct Krassovsky ratio, as shown in Figure (10), the Krassovsky ratio as a function of horizontal wave speed is plotted for an eastward and a northward propagating waves, and that shows the directional dependence of η.

REFERENCES:

Bates, D. R. and M. Nicolet, "Theoretical considerations regarding the altitude of the layer responsible for the nocturnal emission of the sodium D-lines" J. Geophys. Res., 55, 235-239, (1950).

Dodd, J. A., S. J. Lipson and W. A. M. Blumberg, "Formation and vibrationall relaxation of $OH(X^2\Pi_i,v)$ by O_2 and CO_2", J. Chem. Phys., 95, 5752-5762, 1991.

Garcia, R. R. and S. Solomon, "The effects of breaking gravity waves on the dynamics and chemical composition of the mesosphere and lower thermosphere", J. Geophys. Res., 90, 3850-3868, 1985.

Hickey M. P., G. Schubert and R. L. Walterscheid, "Seasonal and latitudinal variations of gravity wave-driven fluctuations in OH airglow", submitted to J. Geophys. Res. Jan. 1992.

Hines, C. O., "Internal atmospheric gravity waves at ionospheric heights", Can. J. Phys., 38, 1441-1481, 1960.

Kaye, J. A., "On the possible role of the reaction $O + HO_2 \rightarrow OH + O_2$ in OH airglow", J. Geophys. Res., 93 (A1), 1988.

Klenerman, D. and I. W. M. Smith, "Infrared chemiluminescence studies using a SISAM spectrometer. Reactions producing vibrationally excited OH", J. Chem. Soc., Faraday Trans. 2, 83, 229-241, (1987).

Krassovsky, V. I., B. P. Potapov, A. I. Seminov, M. V. Shagaev, N. H. Shefov, V. G. Sobolev and T. I. Toroshelidze, "Internal gravity waves near the mesopause and the hydroxyl emission", Ann. Geophys., 33, 347-356, 1977.

Krassovsky, V. I., "Infrasonic variations of OH emission in the upper atmosphere", Ann. Geophys., 28, 739-746, 1972.

Lopez-Moreno, J. J., R. Rodrigo, F. Moreno, M. Lopez-Puertas and A. Molina, "Altitude Distribution of vibrationally excited states of atmospheric hydroxyl at levels v=2 to v=7", Planet. Space Sci., 35 (8), 1029-1038, 1987.

Noxon, J. F., "Effect of internal gravity waves upon night airglow temperatures", Geophys. Res. Lett., 5, 25-27, 1978.

Peterson, A. W. and L. M. Kieffaber, "Infrared photography of OH airglow structures", Nature, 242, 321-322, 1973.

Rensberger, K. J., J. B. Jeffries and D. R. Crosley, "", J. Chem. Phys., 90, 2174, (1989).

Sappey A. and R. Copeland, "Collision dynamics of $OH(X^2\Pi_i, v=12)$, J. Chem. Phys., 93, 1990.

Sivjee, G. G. and R. M. Hamwey, "Temperature and chemistry of the polar mesopause OH", J. Geophys. Res., 92 (A5), 4663-4672, 1987.

Spencer, J. E. and G. P. Glass, "Some reactions of OH(v=1)", Int. J. Chem. Kinet., **9,** 111-122, 1977.

Strobel, D. F., M. E. Summers, R. M. Bevilacqua, M. T. DeLand and M. Allen, "Vertical constituent transport in the mesosphere", J. Geophys. Res., **92 (D6),** 6691-6698, 1987.

Swenson G. R., S. B. Mende and S. P. Geller "Fabry-Perot imaging observations of OH(8-3): rotational temperatures and gravity waves", J. Geophys. Res., **95 (A8),** 12,251-12,263,1990.

Takahashi, H., P. P. Batista, Y. Sahai and B. R. Clemesha, "Atmospheric wave propagation in the mesopause region observed by the OH(8-3) band, NaD, O_2A (8645 Å) band and OI (5577 Å) nightglow emissions", Planet. Space Sci, **33,** 381, 1985.

Tarasick D. W. and G. G. Shepherd, "Effects of gravity waves on complex airglow chemistries,2, OH emission", J. Geophys. Res., **97 (A3),**3195-3208, 1992.

Taylor, M. J., P. J. Espy, D. J. Baker, R. J. Sica, P. C. Neal and W. R. Pendleton Jr., "Simultaneous intensity, temperature and imaging measurements of short period wave structure in the OH nightglow emission", Planet. Space Sci., **39,** 1171-1188, 1991.

Taylor, M. J. and M. J. Hill, "Near infrared imaging of hydroxyl wave structure over an ocean site at low latitudes", Geophys. Res. Lett., **18,** 1333-1336, 1991.

Taylor, M. J., M. A. Hapgood and P. Rothwell, "Observations of gravity wave propagation in the OI (557.7 nm), Na (589.2 nm) and the near infrared OH airglow emissions", Planet. Space Sci., **35,** 413, 1987.

Turnbull, D. N. and R. P. Lowe, "New hydroxyl transition probabilities and their importance in airglow studies", Planet. Space Sci., **37,** 723-738, 1989.

Viereck, R. A. and C. S. Deehr, "On the interaction between gravity waves and the OH Meinel (6-2) and the O_2 atmospheric (0-1) bands in the polar night airglow", J. Geophys Res., **94,** 5397-5404, 1989.

Walterscheid, R. L., G. Schubert and J. M. Straus, "A dynamical-chemical model of wave-driven fluctuations in the OH nightglow", J. Geophys. Res., **92 (A8),** 8775-8780, 1987.

Weinstock, J. "Theory of the interaction of gravity waves with $O_2(^1\Sigma)$ airglow", J. geophys. Res., **83,** 1978

Winick, J. R., "Photochemical processes in the mesosphere and lower thermosphere", in Solar-Terrestrial Physics, edited by R. L. Carovillano and J. M. Forbes, D. Reidel, Hingham, Mass, 677-732, 1983.

Table 1: Rate constants for OH(v) production and loss processes

Rate constants	Value[a]	Comments
k_{1v}	$1.4 \times 10^{-10} \exp(-470/T) \times b(v)$	b(v) is the branching ratio for reaction (1)[b] b(9)=0.48; b(8)=0.27; b(7)=0.17; b(6)=0.08 b(v)=0, v=0-5.
k_2	3.0×10^{-11}	Production of OH(0) only
k_3	$6.0 \times 10^{-34}(300/T)^{2.3}$	Three-body O_3 production
k_4	$5.7 \times 10^{-32}(300/T)^{1.6}$	Three-body HO_2 production
k_5	1.1×10^{-14}	Production of OH(0) only
k_6	$1.6 \times 10^{-12} \exp(-940/T)$	Chemical loss of OH(0) only
k_7	4.8×10^{-11}	Chemical loss of OH(0) only
k_{8v}	$a_8(v) \times 10^{-11}$	$a_8(0)$=3.9; $a_8(1)$=10.5; $a_8(2\text{-}9)$=25.
k_{9v}	$a_9(v) \times 10^{-13}$	$a_9(1)$=1.3; $a_9(2)$=2.7; $a_9(3)$=5.2; $a_9(4)$=8.8; $a_9(5)$=17.; $a_9(6)$=30.; $a_9(7)$=45.96; $a_9(8)$=63.32; $a_9(9)$=81.27
k_{10}	$a_{10}(v) \times 10^{-14}$	$a_{10}(1)$=0.5757; $a_{10}(2)$=1.0; $a_{10}(3)$=1.737; $a_{10}(4)$=3.017; $a_{10}(5)$=5.241; $a_{10}(6)$=9.103; $a_{10}(7)$=15.81; $a_{10}(8)$=27.46; $a_{10}(9)$=47.7
$A_{v,v-n}$		Band-averaged Einstein coefficients from Table-1 of Turnbull & Lowe (1989)

a. T is the temperature in K; reaction rate constant units for a unimolecular reaction is sec^{-1}, for a 2-body reaction are cm^3/sec, and for a 3-body reaction is cm^6/sec.
b. Klenerman and Smith (1987).

357

Applying the perturbation to B in equation (18), and keeping only the first order terms, one gets:

$$\Delta B = Ag \int_z \{C_1 + C_2 - C_3\} \, dz$$

where

$$C_1 = \frac{\Delta N_v(z) \exp\left(-\frac{E}{k_B T(z)}\right)}{Z[T(z)]}$$

$$C_2 = \frac{N_v(z) \frac{E}{k_B T(z)} \frac{\Delta T(z)}{T(z)} \exp\left(\frac{E}{k_B T(z)}\right)}{Z[T(z)]}$$

$$C_3 = \frac{N_v(z) \exp\left(-\frac{E}{k_B T(z)}\right) \Delta Z[T(z)]}{Z^2[T(z)]}$$

Minor Species Profiles, {Day=269, Lat=70N, local midnight}

FIGURE 1

0.1*[O]

[O3]

[HO2]

[H]

Number Density (cm-3)

Altitude (km)

359
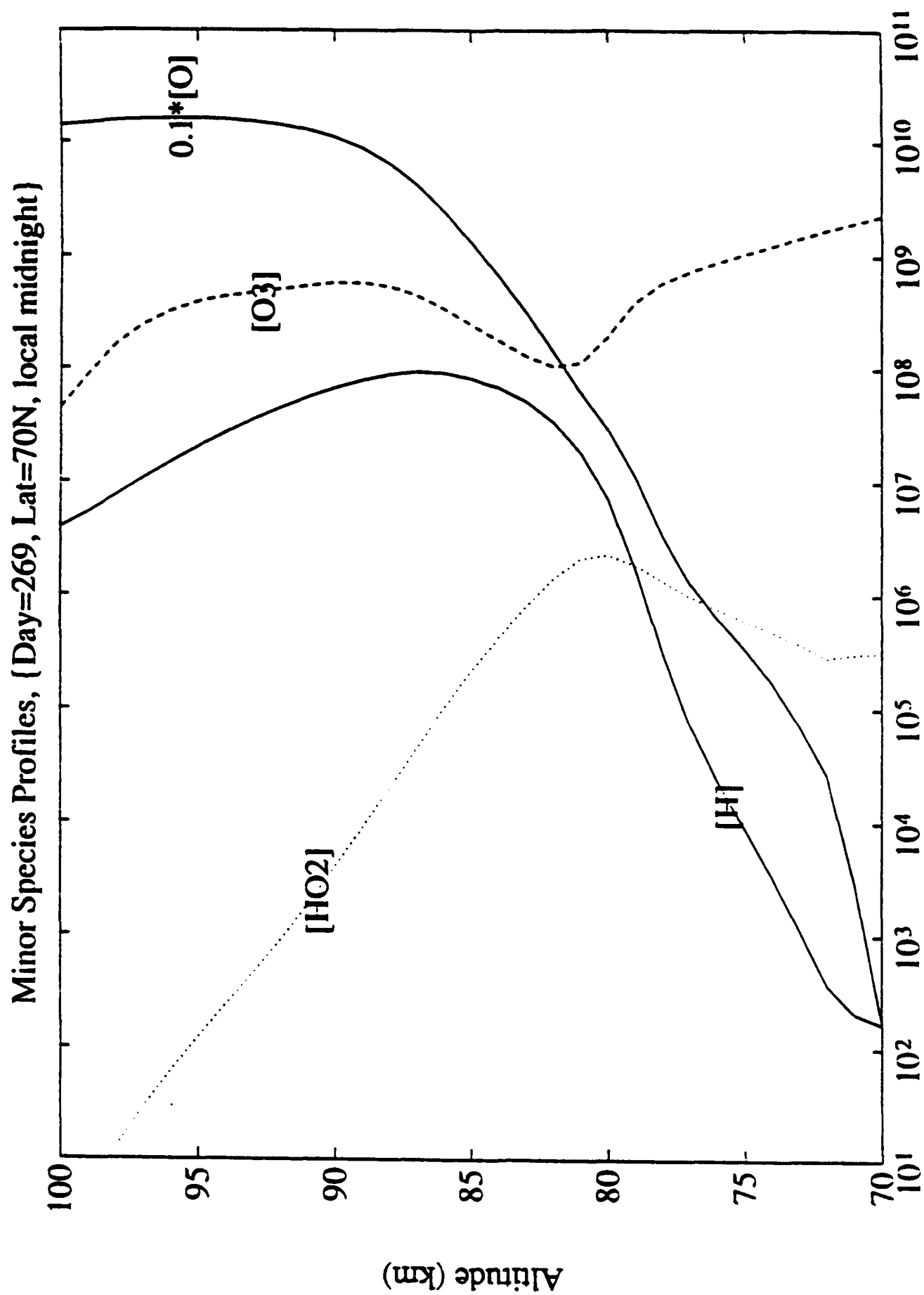
OH vibrational profiles, fast and slow O quenching

slow rate, 2.2E-11

fast rate, 2.5E-10

v = 3

v = 0
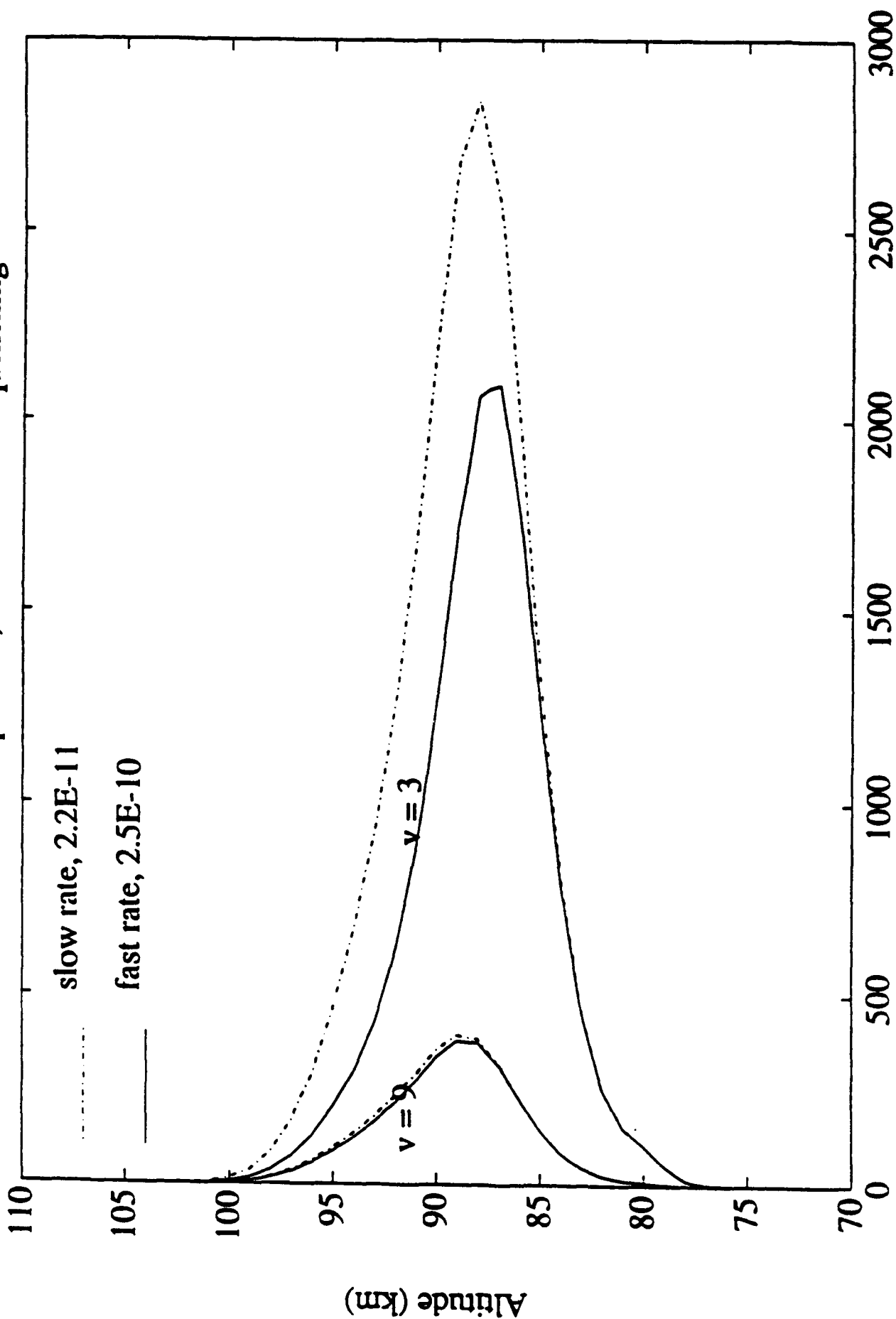
OH(v) number density

FIGURE 2

360

Steady state OH(v) population (v=1-9)

Number Density (cm-3)

FIGURE 3

Altitude in (km)

361
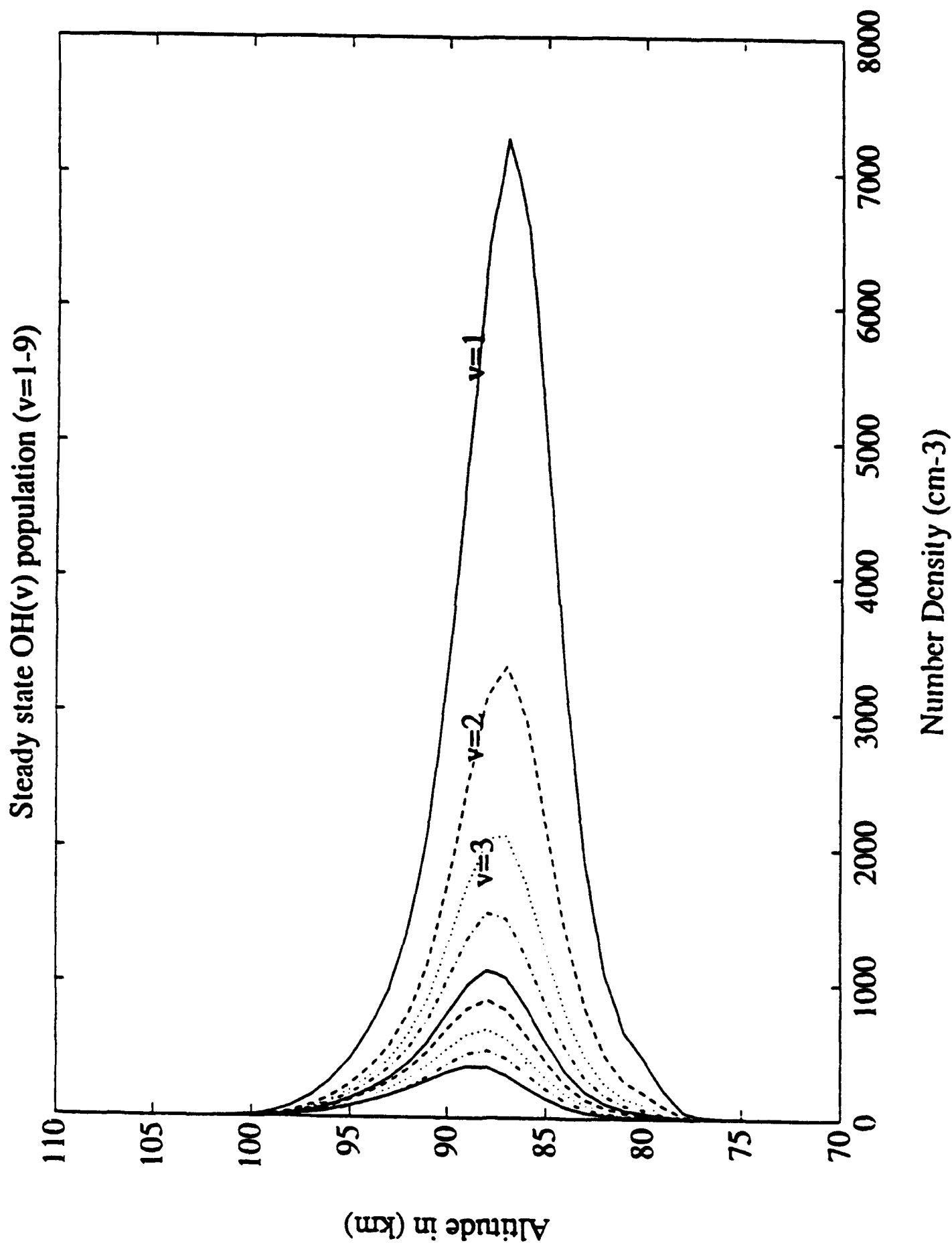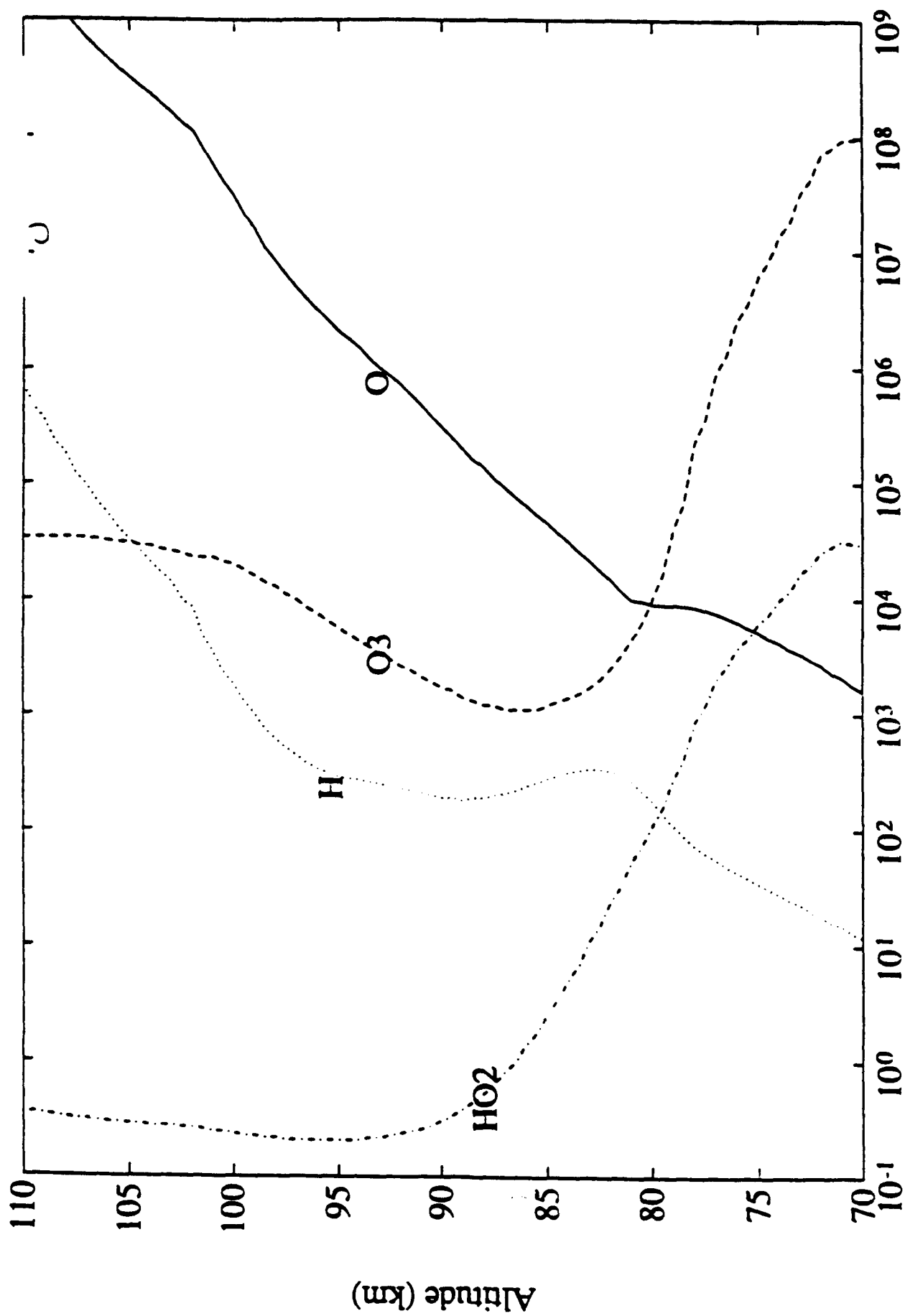
Chemical lifetime (sec)

FIGURE 4

362

OH(6-2), Vphx=70 (m/sec)

DOP
ROT
BWT

Amplitude of eta

Period (sec)

FIGURE 5

363

OH(6-2), Vphx=260 (m/sec)

DOP
ROT
x x x x BWT

Amplitude of eta

Period (sec)

FIGURE 6

364

OH(6-2), Vphx=70 (m/sec)

Phase of eta

Period (sec)

FIGURE 7

DOP
ROT
x x x BWT
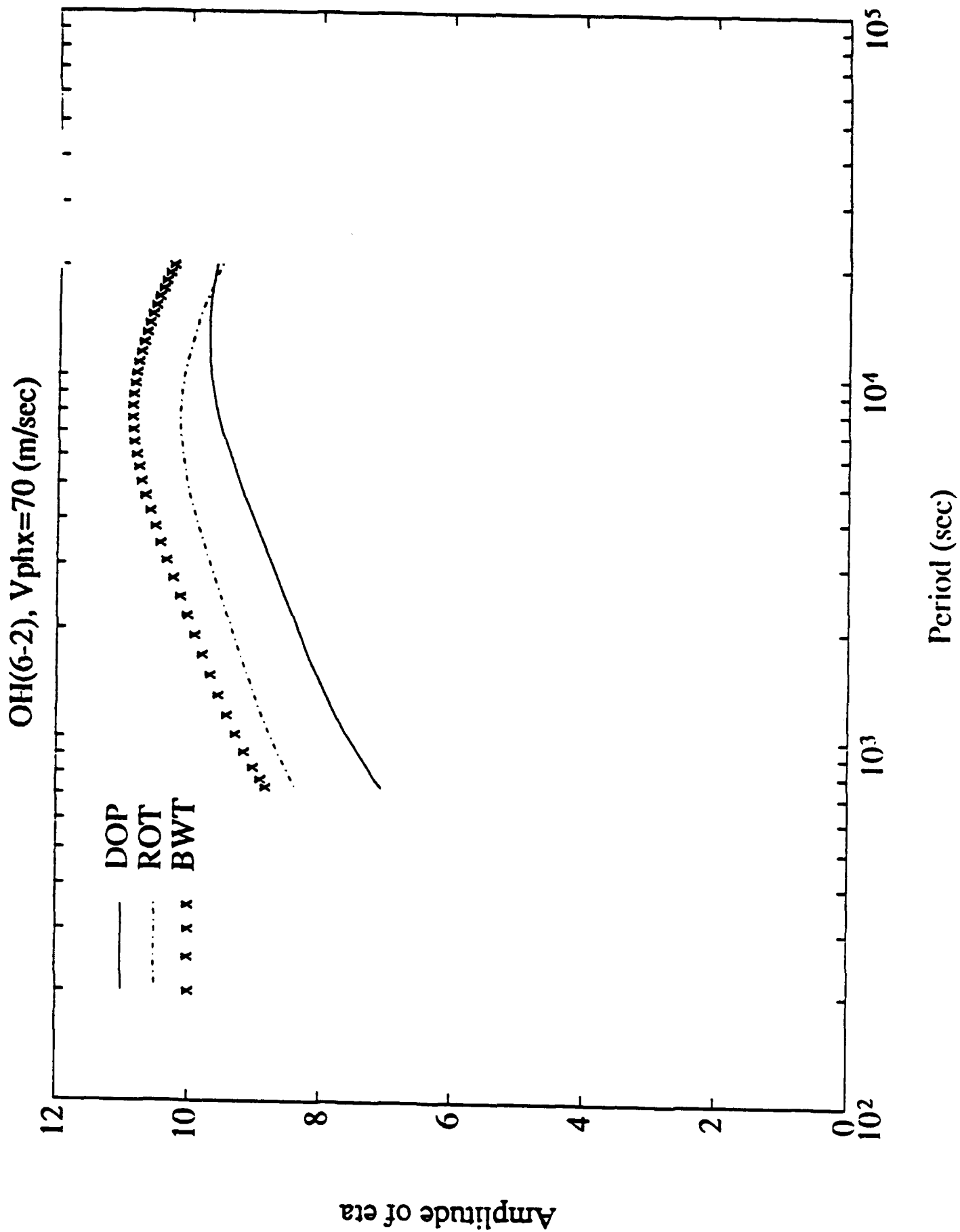
365
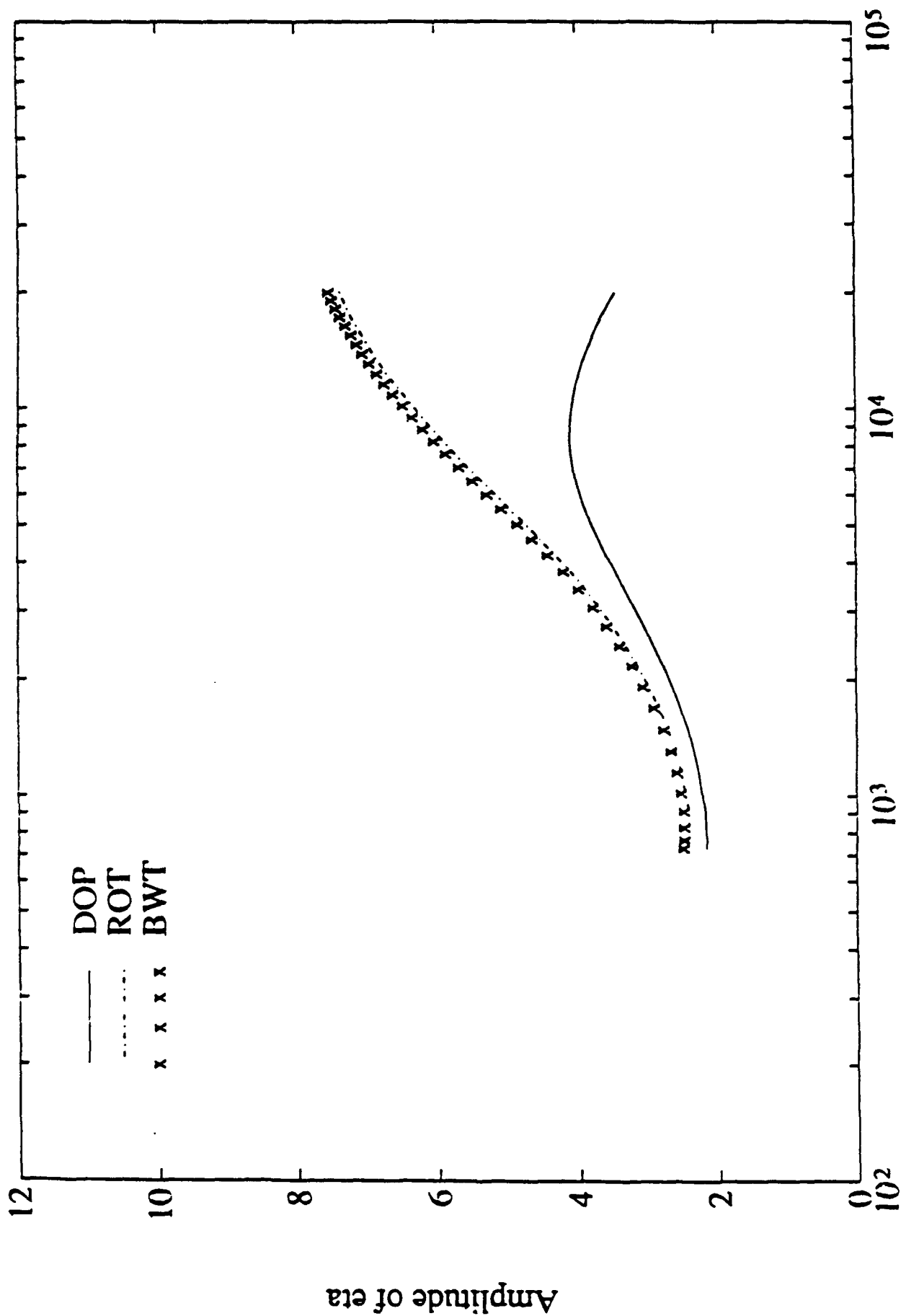
OH(6-2), Vphx=260 (m/sec)
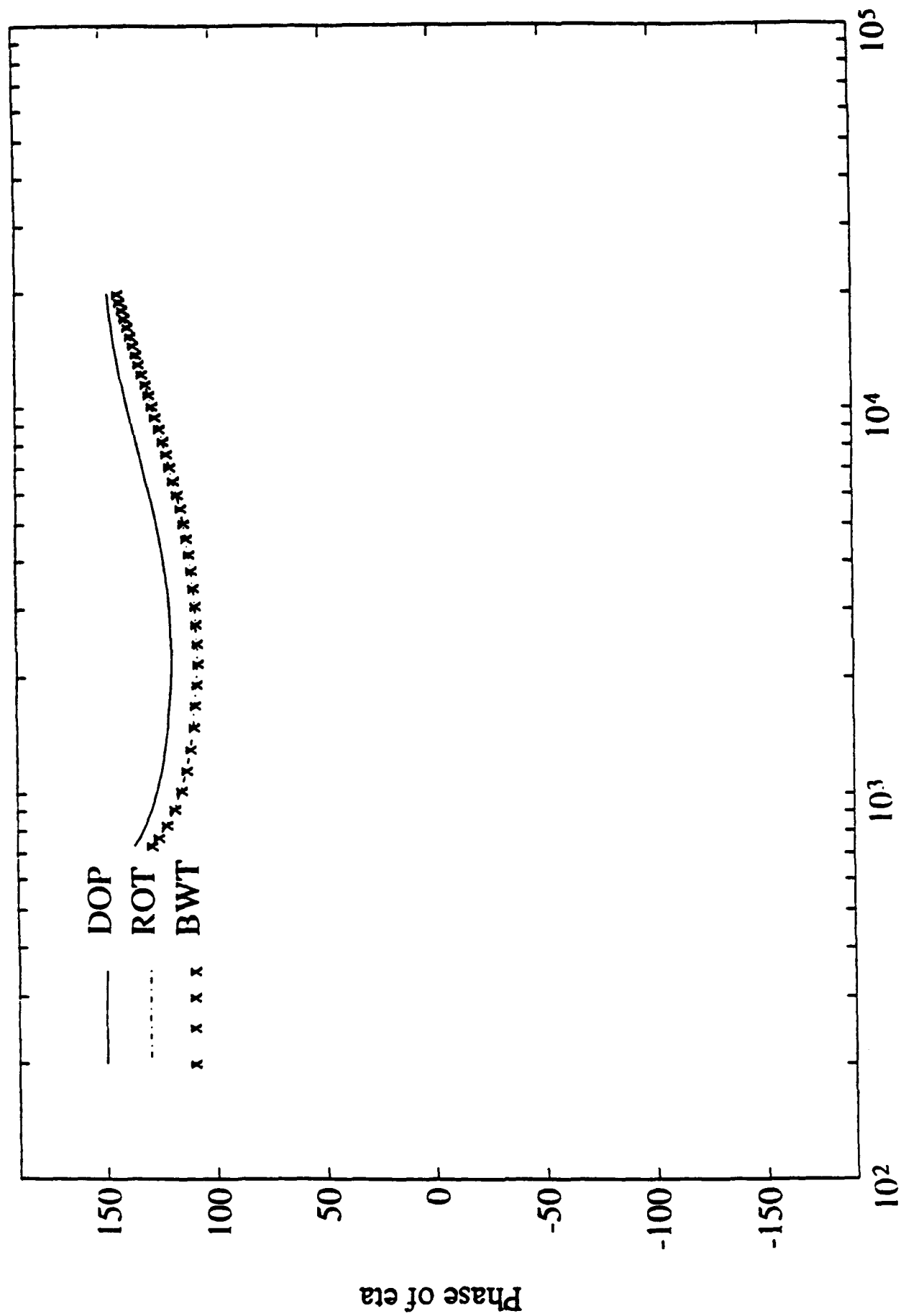
FIGURE 8

Effect of Quenching on OH(v)

.... No Quenching

Vphx= 140 m/sec

OH(6)

OH(9)

OH(3)

Amplitude of Eta

Period (sec)

FIGURE 9

367
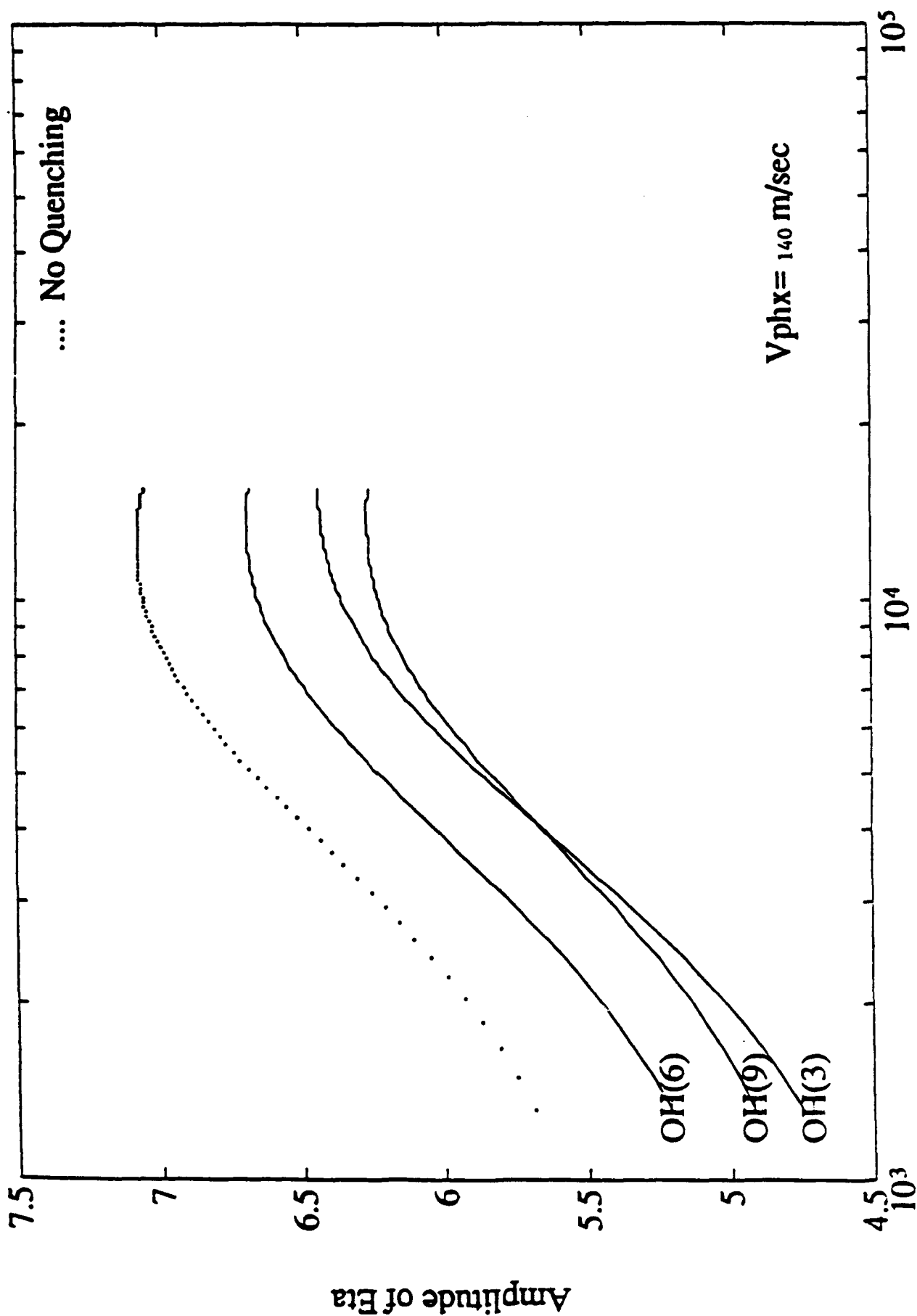
Constant intrinsic Period = 20 min ; OH(3) ; REALISTIC
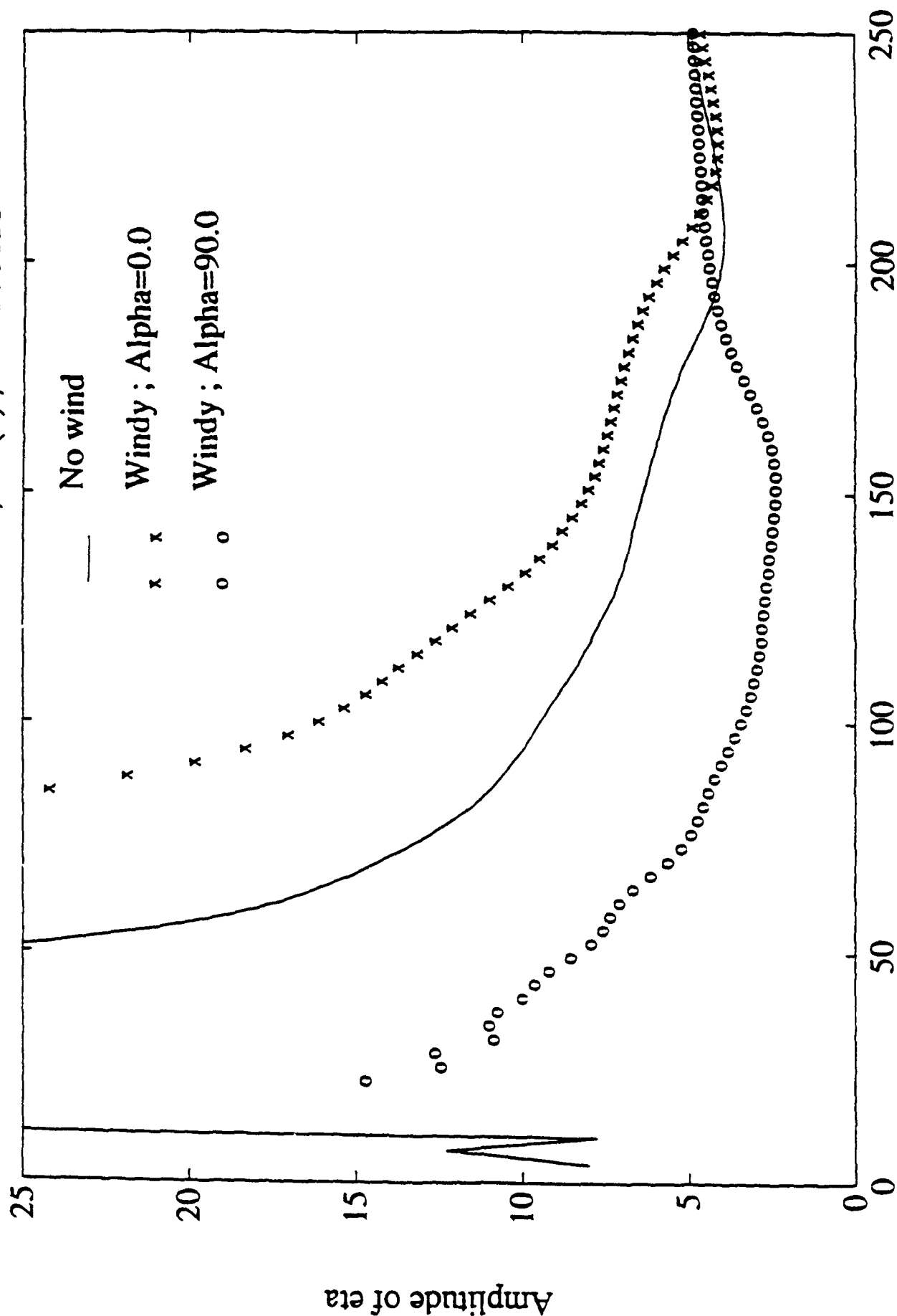
No wind ———

Windy ; Alpha=0.0   x   x

Windy ; Alpha=90.0   o   o

Phase speed (m/sec)

FIGURE 10

Amplitude of eta

368

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

TEST OF COHERENT NEUTRINO DETECTION USING SAPPHIRE CRYSTALS

Prepared by: Dr. Martin P. McHugh

Research Location: Phillips Laboratory,
Geophysics Directorate
Hanscom AFB, MA 01731

AF Research Colleague: Tom Rooney

Contract Number: F19628-86-C-0224

FINAL REPORT ON EFFORT ACTIVITIES

A new temperature control system for the liquid-supported torsion balance (LSTB, on loan from the University of Colorado) was designed, implemented and tested on a seismic pier in the Haskell Observatory. This work is of general interest in improving the performance of the LSTB. This instrument can be used for precision measurements of weak forces, such as tests of the gravitational equivalence principle or use as a gravity gradiometer. In parallel with this work, the LSTB was used for an experiment testing a published result of neutrino detection using sapphire crystals. If the neutrino scattering cross section were large enough, the LSTB would exhibit a clear one day periodic motion due to momentum transfer from solar neutrinos. A null result was obtained and the results were presented at the 1993 Moriond conference (see enclosed paper).

Most recently, work has been done in the testing of equipment for a balloon gravity project. A Global Positoning System (GPS) receiver will be used to track the motion (accelerations) of the balloon in an inertial frame. A strap-down inertial navigation system (INS) will be used to detect all of the accelerations minus the gravitational acceleration. Looking at the difference of these two outputs (using a Kalman filter estimation) the gravity field (averaged over some spatial dimension) can be determined.

At this time a proto-type, ring-laser-gryo INS is being prepared for calibration and testing at Holloman AFB. The INS was developed by Rockwell but was never fully tested. Also, a 4 antenna GPS receiver with attitude output is being tested.

TEST OF COHERENT NEUTRINO DETECTION USING SAPPHIRE CRYSTALS

M. P. McHugh and P. T. Keyser[a]
Air Force Phillips Laboratory/GPEG
Hanscom AFB, MA 01571 USA

ABSTRACT

An experiment to detect solar neutrinos using the method of Weber was undertaken. Two sapphire crystals of about 82 g each, along with compensating lead masses, were placed in a liquid-supported torsion balance (LSTB) in a symmetric configuration. If the sapphire has a sufficiently large coherent scattering cross section (proportional to the *square* of the number of scatterers) then the momentum transfer from solar neutrinos will produce an observable one day period in the angular position of the balance. To the limit of experimental sensitivity, no such effect was observed.

[a]Present address: University of Alberta, Edmonton Alberta T6G2E8 Canada.

INTRODUCTION

Joseph Weber has proposed that neutrino scattering from single crystals can give total cross sections proportional to the square of the number of scattering sites[1-4]. This would produce enhancements of the cross section on the order of Avogadro's number for very modest sized crystals, making a table-top sized neutrino detector possible. The effect differs from ordinary coherent scattering of radiation in that the neutrinos have Compton wavelengths that are short compared to the crystal lattice spacing (for a discussion of long wavelength coherent scattering see ref. 5 page 683). Despite many theoretical arguments against the validity of this claim[6-10], Weber has presented experimental results[3] in apparent agreement with very large scattering cross sections for neutrinos from single crystals of sapphire (Al_2O_3). We decided that an attempt to reproduce his results was merited.

Weber published results for three types of experiments --detection of tritium β-decay neutrinos, nuclear reactor neutrinos, and solar neutrinos; all three utilizing a torsion balance as the detector. We pursued detection of neutrinos from the Sun using essentially the same method as Weber. The technique employs a torsion balance that holds test masses of lead and single-crystal sapphire. If the sapphire has a large scattering cross section and the lead has a negligible cross section then the solar neutrinos will impart momentum to one side of the torsion balance, producing a measurable torque. The torque changes sign as the earth rotates, and the 24 hour periodic motion that results is the solar neutrino signal.

Weber's results were for a torsion balance holding a 26 g sapphire crystal. With 65 days of data averaged together, the amplitude of the 24 h signal implied a force of ~4.6×10^{-6} dynes which he attributed to solar neutrinos. This is an astounding result in that if one assumes that all of the neutrinos that pass through the geometric cross section of this crystal are scattered through a random angle; and that the neutrino flux is that given by the standard solar model (see ref 11), then the maximum force on the crystal would be 5.3×10^{-6} dynes. The curve

designated "predicted signal" in Figure 2 was calculated for our experiment using the same assumptions. The curve is not a sinusoid due to the fact that the geometric cross section of the cylindrically shaped crystals changes throughout the day as the direction of the neutrinos changes. The maximum magnitude of the calculated force is $\sim 2 \times 10^{-5}$ dynes. Other methods for predicting the size of the effect give even larger forces. By simply scaling the results of Weber by the square of the mass for the two 82 g crystals used in this work, we calculate a force of 9×10^{-5} dynes for our torsion balance.

APPARATUS

The apparatus used for this work is a liquid-supported torsion balance (LSTB) that was developed at the Joint Institute for Laboratory Astrophysics (JILA) by Jim Faller and his students[12-14]. The LSTB shown in Figure 1 consists of an aluminum cylinder that floats in water that is kept at its temperature of maximum density (3.98°C). The water provides the support while the spherical and ovoid electrodes on the top of the LSTB are used to provide the centering force and the restoring torque. An optical lever, using four lenses mounted on the lid of the LSTB, provides for the detection of the angular position of the LSTB.

The test masses are two Pb/Al_2O_3 "sandwiches" and six gold-plated Pb cylinders each weighing about 510 g. The crystals themselves weigh 82 g each and are about 4.5 cm in diameter, 1.3 cm in height. The overall height of all eight masses is the same, with the crystal/lead masses having a larger diameter due to their lower density. The symmetric design is used to minimize the effect of gradients in the gravitational field. A calibration of the LSTB sensitivity to applied torques is done in the following way. First the oscillation period is measured for several different electrode voltages. This along with the moment of inertia about the vertical axis is used to calculate the torsion constant (the torque per unit angular displacement). The sensitivity is checked using the gravitational attraction of lead bricks placed near the apparatus. This produces an "order of

magnitude" response and demonstrates that the LSTB is moving freely. The gravitational torque has large uncertainties and cannot be used independently as a calibration.
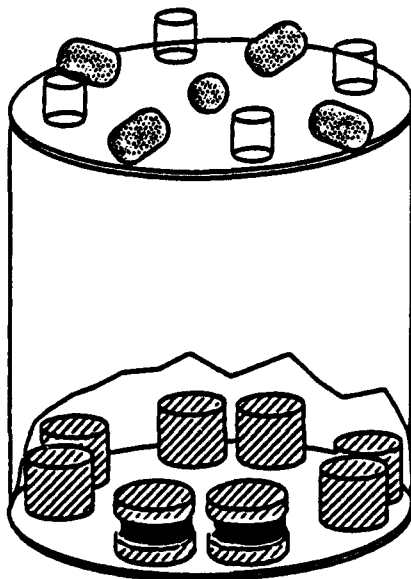
Fig. 1 A cutaway view showing the test masses inside of the LSTB. The sapphire is shown in black. On top are shown the lenses and the electrodes.

DATA

A data run consists of a computer reading multiple channels of an A/D data logger (optical lever output, thermistor resistances, magnetometer output, etc.) at one minute intervals. Ten minute averages are then stored on a disc. The signal is of-the form of a 24 h periodic signal of the proper phase on the optical lever output voltage. This is converted into an angle using a calibration (done just prior to the experiment) and then into a torque using the torsion constant discussed above. Finally this is converted into a force using the length of the moment arm of the crystals about the vertical axis. A graph of the data time series is shown in Fig. 2 along with the predicted signal. A least-squares fit to a function of the form of the expected signal is performed. This functional form is just a cosine of the angle between the zenith and the direction to the Sun. This zenith angle is a fairly complicated function of the time of day and, of course, depends on the time of

year as well. An average of the least-squares fits for three such data runs gives the result of $-3.3\pm2.8\times10^{-8}$ dynes of extra force on the sapphire.
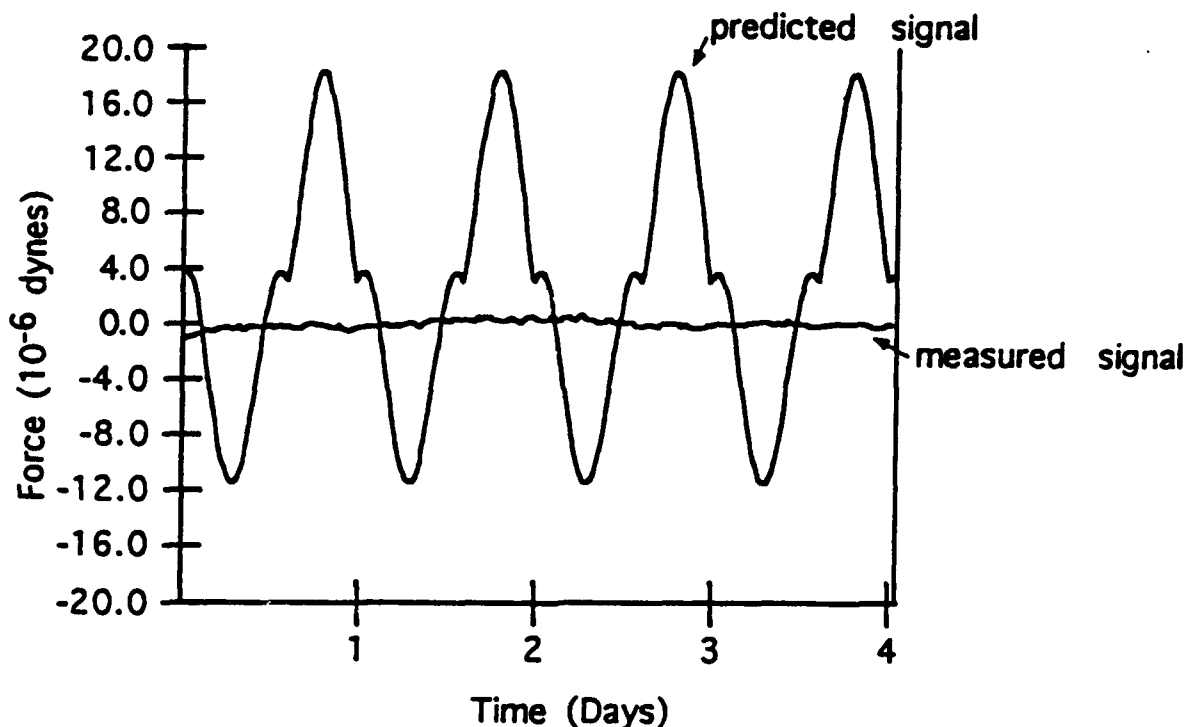
Fig. 2. This graph shows the data from one of the runs (starting at 17:30 EST on 7 Jan. 1993) along with the predicted signal based on the results of Weber. The angular position of the LSTB is converted into a force on the sapphire test masses.

EXPERIMENTAL UNCERTAINTIES

In an experiment of this nature the systematic errors must be dealt with very carefully. The fact that the signal has a period of 24 h presents difficulties in that many effects (e.g. temperature, tilts, motion of people) also occur with this period. The main sources of error considered are, gravity gradients, temperature fluctuations, magnetic field, and tilt. The temperature at various points on the apparatus, the tilt and the magnetic field were all monitored during the data runs in order to look for possible systematic errors.

The gravity multipole couplings are dealt with by first measuring the moments of the float by producing large gradients with nearby lead bricks. Then these moments are used with estimates of the local mass motions to give limits on

375

the torque produced by varying gravtational gradients. This error is estimated to be less than 0.5×10^{-8} dynes. The magnetic field on top of the LSTB tank was monitored with a three-axis flux-gate magnetometer, which combined with measurements of the torques produced by temporary large magnetic fields gave an estimated uncertainty of 0.2×10^{-8} dynes. The tilt was measured with a 2-axis electronic bubble level and in a similar manner the uncertainty was estimated to be 4.0×10^{-8} dynes. The temperature fluctuations at various points on the apparatus were monitored with thermistors, and combined with large artificial thermal gradients to make a worst-case estimate of the uncertainty as 4.0×10^{-7} dynes. The temperature fluctuation contribution dominates the systematic uncertainties, but this seems an overestimate as it is an order of magnitude larger than the "signal". However, temperature effects are very difficult to model and a less conservative estimate seemed unjustified.

An assumed feature of the coherent scattering is that the quality of the crystals is important for the effect. The more nearly perfect the crystals the better the coherence. We had our crystals tested by double crystal X-ray topography, by the same group at NIST that tested Weber's crystals, and they were found to be good single crystals with no extraordinary defects, comparable to his.

CONCLUSION

This experiment shows no observation of enhanced neutrino scattering cross sections for single crystals of sapphire. The extra force on the sapphire was $-0.3\pm4\times10^{-7}$ dynes, consistent with zero. The fraction of the predicted result is -0.002 ± 0.027 compared to 0.86 ± 0.26 observed by Weber. This uncertainty is obtained from his stated uncertainty in the solar neutrino flux[3]. A slightly different approach has also recently obtained a null result[15].

The authors would like to thank Crystal Systems of Salem Massachusetts for providing the high quality sapphire crystals used in this work. Thanks also to Richard Deslattes and Albert Henins of the National Institute of Standards and

Technology in Gaithersburg Maryland for performing the X-ray diffraction topography. We are indebted to Jim Faller for encouraging us to pursue this work.

REFERENCES

1. J. Weber, Found. Phys. **14**, 1185 (1984).

2. J. Weber, Phys. Rev. C **31**, 1468 (1985).

3. J. Weber, Phys. Rev. D **38**, 32 (1988).

4. J. Weber in *New and Exotic Phenomena '90*, proceedings of the Xth Moriond Workshop, Les Arcs, France, 20-27 January 1990 (Editions Frontières, Gif Sur Yvette, 1990) Edited by O. Fackler and J. Tran Thanh Vân, pp. 323-8.

5. J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975) p. 683.

6. G. F. Bertch and Sam M. Austin, Phys Rev. C **34**, 361 (1986).

7. T. H. Ho, Physics Letters **168B**, 295 (1986).

8. Y. Aharonov, F. T. Avignone, III, A. Casher and S. Nussinov, Phys. Rev. Lett. **58**, 1173 (1987).

9. Harry J. Lipkin, Phys. Rev. Lett. **58**, 1176 (1987).

10. M. N. Butler, Phys. Rev. C **35**, 1164 (1987).

11. J. N. Bahcall, *Neutrino Astrophysics* (Cambridge University Press, 1989).

12. G. M. Keiser and J. E. Faller in *Proceedings of the Second Marcel Grossmann Meeting on General Relativity* (North-Holland Publishing Company, 1982) Edited by R. Ruffini, pp. 969-76.

13. P. T. Keyser, Ph.D. thesis, University of Colorado, Boulder, 1986.

14. M. P. McHugh, P.T. Keyser and J. E. Faller in *New and Exotic Phenomena '90*, proceedings of the Xth Moriond Workshop, Les Arcs, France, 20-27 January 1990 (Editions Frontières, Gif Sur Yvette, 1990) Edited by O. Fackler and J. Tran Thanh Vân, pp. 233-6.

15. J. D. Franson and B. C. Jacobs, Phys. Rev. A **46**, 2235 (1992).

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

A STATISTICAL SURVEY OF ELF/VLF WAVES ON CRRES

Prepared by: Dr. Mark Popecki

Research Location: Phillips Laboratory
                                      Geophysics Directorate
                                      Hanscom AFB, MA 01731

AF Research Colleague: William Burke

Contract Number: F19628-86-C-0224

Annual Report

The primary project undertaken in this term was the survey of ELF/VLF electric field measurements aboard the Combined Release and Radiation Effects Satellite (CRRES). Investigation began on an interesting aspect of this survey, which involves the behavior of electromagentic and electrostatic waves when observed by the spacecraft in the same spatial vicinity. Work was also completed on a paper about ground observations of ULF waves at high latitudes. Several presentations were given on these subjects, and they are listed below.

The survey of electric field data required considerable time and care. The object of the survey was to examine average wave spectral densities throughout the region of space covered by the CRRES orbit. Furthermore, the power was to be studied in frequency ranges defined by fractons or multiples of the local electron cyclotron frequency. This choice was made because outside the plasmasphere, which was the focus of the survey, wave power is usually observed to follow the electron cyclotron frequency, and frequency bands so defined would have more physical significance than simple decade bands such as 1-10 kHz. The spatial averaging of spectral densities in the selected frequency bands revealed regions in the magnetosphere where wavegrowth routinely occurs. These regions are selected by the confluence of factors such as electron convection, plasma density and the magnetic field.

The results of the survey are discussed in detail in the attached draft of the survey paper, but they may be summarized here. The whistler mode waves outside the plasmasphere are usually amplified in the midnight-dawn sectors during active times, for frequencies between 30% and 100% of the local electron cyclotron frequency (fce). This is consistent with the conventional idea that wavegrowth arises from plasma sheet electrons that convect sunward within the magnetosphere, around the dawn side of the Earth.

It was also noted that high frequency whistler waves tended to remain closer to the equator than those at low frequencies. The waves between 0.5fce and fce rarely appear above 10 degrees latitude, while those at the lower frequency of 0.3-0.5fce are seen up to 20 degrees latitude. Still lower frequency whistler waves, in the 0.1-0.3fce range, are seen further from the equator and in fact have an intensity maximum off the equator on the dayside. The extent to which the whistlers propagate away from the equator may be related to the orientation of their wavenormals, which can be at larger angles to the magnetic field for lower frequencies. This in turn is important for determining the group velocity and hence the propagation path of the waves. The dayside maximum may be a result of plasmaspheric whistler waves leaking out into the magnetosphere at high latitudes.

Finally, the electrostatic electron cyclotron (ECH) waves that occur between harmonics of the electron cyclotron frequency were also studied. These waves are present at all latitudes reached by the spacecraft, up to 28 degrees magnetic, but strongest in the midnight-dawn sector. They fade below the sensistivity of the electric field instrument on the dusk side, which may be a consequence of electron convection and concomitant loss of energetic electrons.

The ECH waves and the whistler waves appeared to be coupled in such a way that the whistler mode would usually weaken when the ECH wave intensified. This process typically occurred near the equator. An additional feature involved the whistler power above and below half the electron cyclotron frequency, at which a gap is observed. Although both upper and lower bands generally weakened during the ECH intensification, the upper band sometimes remained. The importance of this coupling process lies in the fact that the whistler waves do not appear to freely cross the equator, nor do they propagate more than 20 degrees above it. At the same time, the ECH waves intensify in a

narrow, 2-3 degree latitude range near the equator. The conventional picture of whistler mode amplification may not apply here, as it assumes a gyroresonance between the waves and electrons as the waves repeatedly cross the equator. Instead, the whistler mode waves and ECH waves may be coupled and amplified via the same particle distribution near the equator.

Most of the effort during this term was directed at such problems as the identification and removal of data contaminated by preamplifier saturation or clipping. These conditions would arise from extremely strong natural emissions associated with solar flares and the aurora. A variety of quality checks were conducted on the survey process itself. One of these uncovered an uncertainty of 2-3 degrees between equatorial positions as determined by different magnetic field models. Consequently, latitude bins of 5-10 degrees were used for the survey.

Spectral densities for decade frequency bands were surveyed for an associated project conducted by another researcher. The 0.1-1 kHz band has a physical meaning inside the plasmasphere, where whistler mode waves may propagate far from the region of amplification. They are usually observed to have no relation to the local electron cyclotron frequency, unlike whistler waves outside the plasmasphere. The project studied the possibility of wavegrowth of these plasmaspheric whistler waves, or hiss, by energetic radiation belt electrons. This work was presented at the Fall 1992 Meeting of the American Geophysical Union.

A third project involving Pc1/2 ULF waves was undertaken and completed. It consisted of substantial edits and additions to a paper that had been submitted for publication. This paper analyzed the occurrence of Pc1/2 and Pc1 micropulsations at high latitude ground stations. These pulsations are the ground signatures of 0.1-5 Hz ULF waves that are amplified by ion cyclotron resonances in the magnetosphere. The tendency of these waves to follow the magnetic field makes it possible to deduce their source region by comparing observations at different latitudes on the ground. The paper proposed the location of Pc1/2 amplification and drew conclusions about the role of heavy ions, such as He+, in the amplification and propagation process. This paper was completed and resubmittted to the journal.

Presentations made during the 1 Jan 1992 - 30 June 1993 period:

"ELF/VLF Electric Field Observations on CRRES"; MA Popecki, HJ Singer, C Paranicas, WJ Hughes and RR Anderson, a Space Physics Division Seminar by MA Popecki at the Geophysics Directorate of Phillips Laboratory, 11/12/92, Hanscom AFB, MA

"A Statistical Survey of ELF/VLF Electric Fields by CRRES"; MA Popecki, HJ Singer, C Paranicas, WJ Hughes and RR Anderson, Poster presentation by MA Popecki at the Fall Meeting of the American Geophysical Union, 12/7/92, San Francisco, CA

"The Correlation of Interhemisphere Ground and AMPTE Spacecraft Measurements of Pc1/2 Pulsations"; RL Arnoldy and MA Popecki, Presentation by MA Popecki at the U.S. Antarctic Experimenters' Meeting, 3/18/93, Dartmouth College, Hanover, NH

"Alternating Intensity of Electromagnetic and Electrostatic Waves Outside the Plasmasphere as Observed by CRRES"; MA Popecki, HJ Singer, C Paranicas, WJ Hughes and RR Anderson, Poster presentation by MA Popecki at the Spring Meeting of American Geophysical Union, 5/24/93, Baltimore, MD

"Alternating Intensity of Electromagnetic and Electrostatic Waves Outside the Plasmasphere"; MA Popecki, C Paranicas and RR Anderson, A Space Science Seminar by MA Popecki, 6/10/93, The University of New Hampshire, Durham, NH

A Statistical Survey of ELF/VLF Waves on CRRES

M.A. Popecki and H.J. Singer
Phillips Laboratory, Geophysics Directorate/GPSG Hanscom AFB MA 01731

C. Paranicas and W.J. Hughes
Center for Space Physics, Boston University, Boston MA 02215

R.R. Anderson
Department of Physics and Astronomy, The University of Iowa, Iowa City, IA 52242

Abstract

An extensive survey of ELF/VLF electric field measurements aboard the Combined Release and Radiation Effects Satellite (CRRES) was conducted for the entire mission. Mean spectral densities in four frequency bands were calculated for spatial regions defined by three-hour local time sectors, integer L-shells and three latitude ranges: <5, 5-15 and 15-29 degrees. Spatial coverage included L-shells up to eight, latitudes up to 29 degrees and 3/4 of all local times at apogee. The bands were defined by the local electron cyclotron frequency (fce). Below the cyclotron frequency, bands were selected as $(0.1-0.3)f_{ce}$, $(0.3-0.5)f_{ce}$ and $(0.5-1.0)f_{ce}$ in order to keep track of whistler wave power without individual event selection. Above the cyclotron frequency, a band was defined as $(1-3)f_{ce}$ to follow the first two electron cyclotron harmonic (ECH) components. In the two highest whistler bands, the highest mean spectral densities were on the dawnside for L=5-7, consistent with amplification by substorm-injected electrons. The lowest band displays a dayside off-equator intensity maximum for L-shells of 4-8. In the ECH band above fce, the largest mean intensity is in the midnight-dawn sector, within 5 degrees of the equator. Occasional off-equator intensifications of ECH waves contributed to off-equator midnight mean spectral densities that were only two orders of magnitude less than the largest mean equatorial values.

Introduction

A rich variety of plasma waves may be observed in the ELF/VLF frequency range throughout the plasmasphere and inner magnetosphere. These waves modify particle populations by heating or by pitch angle scattering and consequent precipitation. In the effort to understand the wave-particle interactions that amplify these waves, several spacecraft surveys of this frequency range have been undertaken (Heppner et al., 1992; Russell & Holzer, 1969; Burtis & Helliwell, 1976; Tsurutani & Smith, 1977; Koons & Roeder, 1990; Dunckel & Helliwell, 1969; Muzzio & Angerami, 1972). Many of these have concentrated on the occurrence patterns of whistler mode waves, such as chorus or plasmaspheric hiss. Others have calculated wave-normal angles in the whistler mode and discussed the consequences for growth and propagation (Parrot & LeFeuvre, 1986; Hayakawa et al. 1984; Muto et al., 1987; Sonwalkar & Inan, 1988). Electrostatic electron cyclotron waves (ECH) above the electron cyclotron frequency have also been measured (Paranicas et al., 1992; Roeder & Koons, 1989; Belmont, et al., 1983). Although such

surveys began over 20 years ago, debate continues on the mechanisms responsible for the observed waves.

This study examined the spatially-averaged spectra of waves in the ELF/VLF range, throughout the region of space covered by the Combined Release and Radiation Effects Satellite (CRRES). Electric field measurements were taken for magnetic latitudes of ±28 degrees, magnetic L-values from 1.8 to 8.7, and nearly 18 hours of local time at spacecraft apogee (6.3 RE). CRRES mission space was divided into three hour local time segments, integer L-shells and three latitude ranges. Mean spectral densities ($Volt^2/m^2Hz$) were calculated for each region and are presented below.

The large range in orbital parameters and the nature of some wave-particle interactions suggested the use of a physical parameter, such as the electron gyrofrequency, to organize the data. The plasma frequency would also be an obvious parameter, and may be determined from the upper hybrid resonance when it is indentifiable in the electric field data. The process of calculating density is not yet complete for the entire mission, however.

Using the cyclotron frequency, the bands in the spatially-averaged spectra were chosen so as to pick out physically distinct wave phenomena, such as whistler mode or electrostatic (n+1/2) emissions. Since wave frequency was the only criterion used to distinguish wavemodes, the mean spectra occasionally include broad-band features and transmitters. The frequency range of the sweep frequency receiver (SFR) was 100Hz to 400kHz, and the electron cyclotron frequency ranged from approximately 3 kHz at apogee up to the 400 khz limit close to the Earth.

Instrumentation, Orbit, Sample E-field Data and the Survey

The Orbit

The CRRES orbit was approximately 10 hours and the orbital plane was inclined 18 degrees with respect to the geographic equator. Precession moved the apogee from 0745 through midnight to 1343 local time (LT). Useful data for this study began at an apogee LT of 0708. Apogee was at 6.3Re, and at the highest magnetic latitude, CRRES reached an L-value of 8.7. Perigee for the spacecraft was 1.06Re. Frequency constraints of the SFR/dipole system limited the minimum L-value to 1.8 for this study.

The combination of the 18 degree inclination, precession and the 10 hour period allowed CRRES to sample a 22 degree range of magnetic latitudes every three orbits. When apogee was the highest above the geographic equator, magnetic latitudes at apogee spanned 7-29 degrees. When the apogee had precessed to the intersection of the equatorial and orbital planes, the latitude range at apogee was ±11 degrees. Equator crossings took place at other radial distances on every orbit. The absolute value of latitude was used to improve sampling in this study. The time spent by CRRES in each spatial bin of this study, as well as the number of visits to those regions, are shown in Figure 2.

The spin axis was pointed approximately toward the Sun. The 100 m wire dipole from which wave measurements were made rotated in the plane perpendicular to the spin axis with a period of approximately 30 seconds.

Instrumentation

The electric field measurements for this survey were made with a 100m dipole and a 128 channel sweep frequency receiver (SFR). These are described in Anderson & Gurnett (1992). The SFR channels ranged from 100 Hz to 400 kHz, in four bands of 32 logarithmically spaced frequency steps. The frequency ranges were: 100-810 Hz, 810-6400 Hz, 6.4-51.7 kHz and 51.7-400 kHz. The four bands had different channel bandwidths (7, 56, 448 and 3600 Hz) and sampling rates (1, 2, 4 and 4 channels/sec). Each band had a dynamic range of approximately 100 dB.

The electron cyclotron frequency was calculated from magnetic field measurements by a fluxgate magnetometer. The magnetometer is described in detail by Singer et al. (1992).

A Sample Orbit

A sample orbit is shown in Figure 1. An orbit consists of approximately 10 hours of data. Ephemeris values and UT are shown on the horizontal axis and frequency (100 Hz-400 kHz) on the vertical axis. Also shown in the figure are magnetic equatorial and meridian projections of the orbit. Apogee occurred in the midnight sector, within 10 degrees of the magnetic equator. Latitude was calcuated with the Olson- Pfitzer (1977) field model. The local electron cyclotron frequency (f_{ce}) appears as a solid line, minimizing near 1000 UT. Also shown are lines representing $0.1f_{ce}$, $0.3f_{ce}$, $0.5f_{ce}$, f_{ce} and $3f_{ce}$.

This orbit was chosen as an example because nearly every type of wave seen in the mission appears here. Apogee occurs near the center of the plot. The horizontal striations are electrostatic emissions at approximately (n+1/2) harmonics of the local cyclotron frequency. They intensify at the equator, which occurs near 0945 in this orbit.

The plasmapause is indicated by the steep drop in the upper hybrid frequency at 0845 and 1430. The upper hybrid frequency is approximately equal to the plasma frequency over most orbits because the cyclotron frequency is usually much less than the plasma frequency. Outside the plasmasphere, the upper hybrid resonance can be difficult to locate because of the broad-band features (for example from 1200-1240), continuum radiation or the highest (n+1/2) harmonic.

The intense emission after 1400 UT, below 10 kHz, is plasmaspheric hiss. It is contained inside the plasmapause. These are whistler mode electromagnetic emissions and they have been discussed by Lyons & Thorne (1970), Thorne (1973), Huang, et al. (1983) and others. Plasmaspheric hiss often appears at frequencies below the local lower hybrid frequency. The lower cutoff for this study was set at $0.1f_{ce}$, which is above the lower hybrid frequency, but is still compatible with the lower limit of the SFR in the weakest magnetic fields observed by CRRES.

A Type III radio burst occurred at 1100 and approximately 100 kHz in this orbit. Auroral kilometric radiation (AKR) was observed at 0830 above 200 kHz. Intense type III bursts and AKR can produce artificial broad-band signatures in the top two 32 channel bands. These signatures were removed from the data set before analysis.

The constant frequency lines below the cyclotron frequency at 1500 UT are transmitters. They appear mostly below L=3.5 where they are damped to the local electrons at the local electron cyclotron frequency. They can be stronger than any natural waves observed at the same time. They have not been removed from the data set, however the features discussed in this paper are primarily beyond the L-shells in which they are seen.

383

Finally, the strong emissions below the cyclotron frequency at 0900 are whistler mode waves known as chorus. A gap is usually present at half the gyrofrequency. The presence of the gap has been discussed by previous authors (Tsurutani & Smith, 1974; Burtis & Helliwell, 1976; Maeda, 1976; Koons & Roeder, 1990). These emissions often consist of burst-like features, separated by approximately one second. The chorus elements below the gap are usually unrelated to those above, although they sometimes may cross the gap (Burtis & Helliwell, 1976). The time resolution in the CRRES SFR is insufficient to see the chorus elements. Other chorus emissions are structureless (Tsurutani & Smith, 1974). The wave-normal angles of the chorus above and below the gap at half the cyclotron frequency have been discussed by Hayakawa et al. (1984), Muto et al. (1987) and Sonwalkar & Inan (1988). Wave-normal angles were not obtainable for whistler waves aboard CRRES, however they may explain the spatial intensity patterns in this study because of their importance to wave growth and propagation.

The Survey Process

The goal of the survey was to calculate the average ELF/VLF spectral densities for selected frequency bands in the regions of space visited by CRRES. The bands were chosen so as to keep track of physically separate wave phenomena where possible. The highest frequency band covered the frequency range $(1-3)f_{ce}$, which consists mostly of electrostatic $(n+1/2)f_{ce}$ harmonics (ECH). The next three bands were below this and contain mostly whistler mode waves. These might be chorus and hiss, however the SFR did not have sufficient time resolution to distinguish the two. From highest to lowest, the band were: $(0.5-1.0)f_{ce}$, $(0.3-0.5)f_{ce}$ and $(0.1-0.3)f_{ce}$. The break at $0.5f_{ce}$ was chosen because the gap that is often observed in whistler wave power occurs at that frequency. The $0.3f_{ce}$ break is close to the lower bound of the lower chorus band (Burtis & Helliwell, 1976; Tsurutani & Smith, 1974).

A 128 channel sweep from the SFR is shown from orbit 515 at 09:09:59 UT in Figure 3. The electron cyclotron frequency is noted, and the channels within each of the four bands are separated by vertical lines. In all four bands, the spectral features typically have a wider bandwidth than the separation between individual channels.

Within each band, spectral densities (v^2/m^2Hz) from each channel were summed. Every eight consecutive sums were then averaged, representing 65.5 seconds of data. The sums were divided by the number of channels in the band to give an average power spectral density for the band. Mean power spectral densities for each band in orbit 515 are shown in Figure 4. Spatial coordinates, such as L-shell crossings, latitudes and local times are identified. The ECH emissions in the $(1-3)f_{ce}$ band reach a level of sum of 5×10^{-10} v^2/m^2Hz within 5 degrees of the equator, just after the power in the upper two whistler bands (covering $(0.3-1.0)f_{ce}$) subsides. The effect of the transmitters may be seen in the two upper whistler bands, particularly near 1530 UT, where the spectral densities reach $8 \times 10^{-16} v^2/m^2Hz$ before falling rapidly.

The logs of the 65.5 second mean spectral densities were then spatially binned and averaged by integer values of L-shell, 3 hour increments of local time and three latitude ranges: <5; >5,<=15; and >15, <29 degrees. In order to estimate confidence in these regional mean values, also recorded were the numbers of visits to each region, the amount of time spent there, and the mean spectral density per visit.

The $(n+1/2)f_{ce}$ electrostatic emissions are clearly organized by the local electron cyclotron frequency. On the other hand, the whistler mode spectrum may be sensitive to both the equatorial (Dunckel & Helliwell, 1969) and local cyclotron frequency. In the whistler mode bands, the spatial averages were calculated separately with respect to the local and equatorial cyclotron frequency, using the Olson-Pfitzer 1977 model for the latter case. The results are grossly similar.

Results for the Whistler Mode Bands (below f_{ce})

The mean logs of spectral densities for the bands below the local electron cyclotron frequency are presented in Figures 5 and 6. The averages are shown in local time/L-value polar plots, one for each frequency band and latitude range. For example, the plot for the $(0.3-0.5)f_{ce}$ band at $<5,>15$ degrees presents the mean spectral densities from all measurements in that frequency range, while the spacecraft was in that latitude range. Concentric circles represent L- values, with the inner circle as L=1. White spaces indicate a lack of data. Either CRRES did not visit a region or there were fewer than 5 visits there, or else that the average visit to a region was less than 10 minutes.

Figure 5 has spectral densities for Kp > 1o, and Figure 6 has them for quiet times (Kp<=1o). The $(0.5-1.0)f_{ce}$ band has the largest mean values in the dawn sector for L=5-6, in the ±5 degree latitude range and for Kp>1o. This band typically represents the whistler mode waves above the gap that usually appears at $0.5f_{ce}$. The dawn sector mean decreases with increasing latitude, indicating that these waves are confined near the equator. In fact, the waves in this band are usually seen in the CRRES data no higher than 10 degrees. This band appears to be associated with geomagnetic activity, since the dawn sector equatorial mean is reduced during quiet times (Kp<=1o, Figure 6).

The middle whistler band, $(0.3-0.5)f_{ce}$, also shows a dawnside equatorial mean intensity maximum that reduces with increasing latitude. At the highest latitudes, the mean spectral densities for this band are comparable to those in the $(0.5-1.0)f_{ce}$ band, but in the middle latitude range, they are greater for the $(0.3-0.5)f_{ce}$ band. This indicates that the $(0.3-0.5)f_{ce}$ band emissions extend further above the equator than the highest whistler band.

The lowest frequency band, $(0.1-0.3)f_{ce}$, exhibits a high- latitude (15-29 degrees) dayside maximum mean spectral density for Kp > 1o (Figure 5). A remnant persists during low Kp (<=1o) in the 12-18 LT sector.

An example of the kind of data that contributes to the high- latitude dayside maximum spectral density is shown in Figure 7. This orbit had an apogee near 1330 LT, within 10 degrees of the equator. On the outbound leg, CRRES was at high negative latitudes in the prenoon sector as it approached apogee. Whistler mode power was observed at the high latitudes (0600-1130 UT), however it became less intense in the $(0.1-0.3)f_{ce}$ band as CRRES approached the equator. Kp values for this orbit ranged from 3- to 4+.

During the following orbit, although CRRES was within 7 degrees of the equator at apogee, and Kp levels were between 4o and 6o, less intense whistler activity was observed. These examples are not fully representative of the statistical study, but this type of observation was common when the CRRES apogee was on the dayside, which occurred at the beginning and end of the mission.

Visit-averaged Spectral Densities

The mean spectral densities per visit by CRRES to selected regions is shown in Figures 8 and 9. Also shown are the number of visits by CRRES to that region and the amount of time spent there. This presentation shows the distribution of measurements on the dawnside and the dayside that lead to the whistler band features in those sectors.

The Dawnside

Figure 8 has the spectral densities for all three whistler bands for the dawn sector region defined by: 03-06 LT, L=5-6, Kp>1o and both low (£5) and high (15-29) latitudes. The spacecraft visited this region 75 times at low latitudes and 67 times at high latitudes. Visits were usually 20-40 minutes near the equator and 50-80 minutes above 15 degrees.

Both the (0.5-1.0)fce and (0.3-0.5)fce bands have greater mean spectral densities near the equator than above 15 degrees. On the other hand, there is little difference between high and low latitudes for the lowest band, (0.3-0.5)f$_{ce}$, in this local time sector.

The Dayside

The dayside 12-15 LT sector is shown in Figure 9, for L=5-6 and Kp>1o. The spacecraft visited the near-equator region 55 times, most commonly for 70-80 minutes each. The 15-29 degree latitude range was visited 101 times, typically for 30-60 minutes each.

In the highest whistler band, the largest spectral densities are again observed near the equator. In the middle band, values as high as 10^{-10} are seen in both latitude ranges. In contrast to these, the largest spectral densities in the lowest band, (0.1-0.3)fce, are most likely to be seen in the 15-29 degree latitude range instead of the near-equatorial region.

Spectral Density per Visit vs. Kp and Local Time

Spectral densities for all local times, one latitude and L-shell range, and one whistler band are plotted vs. Kp in Figure 10. The data are for latitudes of 5 degrees or less, L=5-6 and frequency band (0.1-0.3)f$_{ce}$. The largest mean spectral densities per visit occurred in the midnight (21-06 LT) sectors.

In dawnside local time sectors, there is an approximately proportional relationship between Kp and mean spectral density. This proportionality vanishes, however, for 12-21 LT. In this sector, there seems to be no relationship between the mean spectral density per visit and Kp at the time of measurement.

The lack of proportionality between Kp and spectral densities from 12-21 LT may arise from the drift time of the electrons that amplify these waves. In the midnight-predawn sector, the Kp values at the time of measurement could be simultaneous with the injection of the electrons that amplify the observed waves. On the dayside,these two may no longer be simultaneous due to the electron drift time and the three-hour Kp binsize. The lack of Kp-spectral density proportionality after 12 LT suggests that electrons may drift there in about 1.5 hours or more. An upper limit of electron energy may be roughly estimated from this time limit. A complete drift time of 3 hours for an equatorially mirroring electron in a dipole field at L=5.5 would be would be approximately 44 keV (estimated from Davidson,

1977), which is an energy suitable for amplification of whistler mode waves below $0.5f_{ce}$ (Inan et al., 1992).

Survey Results for the (1-3)Fce Band

Figure 11 shows the mean logs of spectral density for each of the three latitude ranges, at all local times and L- shells for the (1-3)f_{ce} band. The means were calculated for low Kp ($<1o$) and high Kp ($>1o$).

The largest mean spectral densities occurred within 5 degrees of the magnetic equator, in the 21-09 (midnight/dawn) sector, for L>=4. The lowest mean values within 5 degrees of the equator were in the 15-18 LT sector. Between 15 and 29 degrees, the lowest occurred in the 12-18 sector, a somewhat broader LT range

Although the (n+1/2)f_{ce} emissions tend to intensify near the equator, the spectral density values in Figure 11 are lower than these intensifications because they are averages taken over a ±5 degree range across the equator. The intensifications are not always present through the entire range, either because the wavepower is more closely confined than 5 degrees or because of uncertainty in the calculation of the equatorial position.

Moreover, since the intensifications do occur mostly near the equator, the mean spectral densities at higher latitudes, above 5 degrees, are primarily descriptive of the banded emissions that are usually present, such as those in Figure 1, orbit 515, from 1300-1400 UT. It is these banded emissions that are weaker in the 15-18 LT sector. Occasionally, they do not even appear in that sector above 10^{-17} v^2/m^2Hz, which is at the sensitivity limit of the instrument.

For some insight into how large the spectral densities became, histograms of spectral densities per visit to the 00-03 LT sector at L=6-7 and Kp>1o are presented in Figure 12. The largest visit-averaged spectral densities were 10^{-10} v^2/m^2Hz. Within 5 degrees of the equator on one orbit in this sector, the CRRES survey detected the largest average amplitude per visit of 0.6 mV/m.

At high spacecraft latitudes (15-29 degrees), ECH emissions occasionally intensify such that the spectral density in the midnight L=5-7 regions could reach $10^{-9.6}$ v^2/m^2Hz. Figure 13 shows an example of an ECH intensification away from the equator, in orbit 0635 at 1700 UT. The amplitude at 16:57 UT for the lowest ECH harmonic, (1+1/2)f_{ce}, was 0.3 mV/m.

Also contributing to the (1-3)fce band were impulsive broadband features of the type seen at 1230 UT in orbit 0515 (Figure 1). During magnetically disturbed times, when the local magnetic field and electron cyclotron frequency is reduced, the broadband power above the electron cyclotron frequency may be observed by the spectrum analyzer/search coil antenna on CRRES. A magnetic component is sometimes observed above the cyclotron frequency for the broadband features.

Figure 11 shows mean logs for quiet Kp levels (Kp<=1o). Mean spectral densities did not exceed $2x10^{-15}$ v^2/m^2Hz near the equator and $9x10^{-17}$ v^2/m^2Hz above 15 degrees. On the nightside, the most intense region is further from the Earth than during Kp>1o.

Discussion

The Dayside, Off-equator Intensity Maximum

The dayside, off-equator intensity maximum in the $(0.1-0.3)f_{ce}$ band is a prominent feature of the whistler data. This phenomenon appears clearly in the lowest frequency whistler band and may extend below it, as orbit 1049 (Figure 7) suggests. Orbits such as 1049 and the isolation of this effect to the whistler band below 0.3 f_{ce} suggest that the power is at least roughly associated with the local electron cyclotron frequency.

The high latitude (>15 degrees) whistler power was in some cases observed to decrease in frequency as the spacecraft moved outward, over distances of as much as 2Re, before fading. Orbit 1049 in Figure 7 is an example of this. These decreases were not clearly controlled by either the local cyclotron frequency or the equatorial cyclotron frequency for the field line of the spacecraft. Nevertheless, it is possible that the waves might be amplifed at the equator and then propagate to the spacecraft, although not necessarily on the field line of amplification.

If the whistlers are amplified at the equator, however, it is remarkable that the observed disparity in spectral densities exists between high and low latitude measurements. Figure 9 shows that CRRES visited the 12-15 LT, L=5-6 sector 56 times for latitudes <=5 degrees, and 101 times above 15 degrees. Visits to latitudes above 15 degrees on the dayside were separated from visits to the equatorial region by at least 10 hours, since the spacecraft inbound leg is at different local times and latitude bins than the outbound leg. In the $(0.1-0.3)f_{ce}$ band, CRRES observed mean spectral densities per visit of 1^{-9} and 10^{-8} a total of 48 times for latitudes above 15 degrees, compared to twice near the equator (<= 5 degrees). The high-latitude distribution is peaked at 10^{-8}, while the near-equatorial distribution is peaked near 10^{-11} v^2/m^2Hz. This could occur if equatorial visits happened to be during times when wave power was weak at all latitudes on the dayside.

As a test of this possibility, the distribution of Kp values at each visit to the dayside regions was compared to the mission Kp distribution with a chisquare test, to see if samples were made under a variety of magnetospheric condtions. Kp values for the 12-15 LT sector at high latitudes and L=4-7 were in good agreement with the mission Kp distribution, with at least 85% probability of exceeding chisquare. Near the equator, the probability was 40% for the L=6-7 region, and at least 60% for L=4-6. This suggests that the dayside regions were sampled under a range of conditions similar to those during the mission as a whole. CRRES should have found itself on the dayside equator during times of strong whistler power just as often as it did above 15 degrees.

It is also possible that the most intense whistlers were amplified on the equator further away than 6.3Re, the CRRES equatorial apogee, and arrived at CRRES off-equator positions via unguided propagation. The equatorial CRRES positions might not be as accessible for such a source. It should be noted that the mean spectral density at the highest dayside latitudes for L=7-8 is less than for L=4-7. At L=7-8, CRRES should be closest to a source region in the high latitude magnetic field minima.

Thorne et al. (1973) and references therein suggested that some whistler waves may leak from inside the plasmasphere from high latitudes on the dayside. The plasmapause density gradient is expected to be weaker there than on the nightside because of the outflow of ionospheric plasma. Thorne, et al. also expected Landau damping of the escaped whistler

waves by ring current protons. Outside the plasmasphere, this power might still be below the local electron cyclotron frequency, since whistler power inside the plasmasphere is quite often observed to be peaked at under 1 kHz and is usually below the local lower hybrid frequency (Lyons et al., 1972 and references therein).

Russell et al., (1969) observed ELF hiss (100-1000 Hz) on OGO 3. They found on the dayside a larger occurrence of steady hiss between 30-50 degrees latitude and beyond L=5 than on the equator.

Tsurutani & Smith (1977) used an OGO 5 search coil experiment to classify chorus (whistler mode waves) as either equatorial or high latitude. The equatorial chorus was found in the postmidnight and post-dawn sectors. High-latitude chorus, found at magnetic altitudes above 15 degrees, occurred mainly on the dayside, between 08 and 16 LT. The high latitude source was within 1-2 Re of the magnetopause. The authors concluded that this type of chorus might be amplified where the magnetic field minimizes, at latitudes of 20-50 degrees on the dayside, due to compression by the solar wind (Roederer, 1970). These minima in B are lower than at the equator on the same field line. The high latitude chorus observed in their study had peak occurrences with frequencies of either less than $0.15f_{ce}$, or else of $(0.25-0.30)f_{ce}$, where f_{ce} is the local electron cyclotron frequency. On CRRES, such waves would be observed at lower fractions of the local electron cyclotron frequency.

Parrot and LeFeuvre (1986) studied ELF hiss with GEOS-1 and found that outside the plasmapause, there is a peak occurrence for 11-13 LT and 20-30 degrees magnetic latitude. They calculated propagation directions and concluded that just outside the plasmapause, the observed ELF could be either leakage from the plasmasphere at high latitudes, or a source region at latitudes above 20 degrees. Far from the plasmapause, they concluded that an equatorial source region existed for ELF hiss.

The Localization of Whistlers Near $f_{ce}/2$ to the Equatorial Region

The $(0.5-1.0)f_{ce}$ band and the $(0.3-0.5)f_{ce}$ band cover the whistler waves above and below the gap that often occurs at $0.5f_{ce}$. In the $(0.5-1.0)f_{ce}$ band, the mean spectral densities peak in the 03-06 LT sector on the equator. In this sector, the upper band spectral densities fall off faster with latitude than those of the middle band, $(0.3-0.5)f_{ce}$. For L=5-6 in this LT sector, their ratio is 0.7 for latitudes of five degrees or less, 0.01 for 5-15 degrees and 0.23 for 15-29 degrees. At the highest latitude range, the two bands have comparable values.

In the same sector and L-shell range, the lowest band, $(0.1-0.3)f_{ce}$, had a mean spectral density of about one-third of the middle band, $(0.3-0.5)f_{ce}$, within five degrees of the equator. However, above 15 degrees, the ratio reversed, and the lowest band had a mean value two orders of magnitude greater than that of the middle band. The lowest frequency band has more power further from the equator than the higher frequency bands. This may again point to an off-equator source or else to low frequency wavepower that more closely follows the magnetic field away from the equator than the waves in the upper bands.

Whistlers in the upper band were not usually seen above approximately 10 degrees, however those in the lower band would appear at higher latitudes. Hayakawa et al. (1984) calculated wave normal directions of chorus emissions above and below the gap at GEOS 2 (L=6.6). They found that above the gap, the chorus wavenormal angles were close to the resonance cone and the waves were consequently quasielectrostatic. In that case, their

phase speed would be small and they might Landau dampen before propagating far from the equator. The group velocity of these quasielectrostatic waves might also keep them in the equatorial plane, propagating at large angles to the magnetic field. Below the gap, wavenormal angles were found to have a range of values below the resonance cone.

The gap may sometimes be seen at higher latitudes, though. Muto et al. (1987) published two events from GEOS 1 in which a gap in whistler power was observed at 17 and at 26 degrees latitude on the dayside. L values were 7.6 and 8, respectively. These regions were not covered well by CRRES due to the ending of the mission. They concluded that these waves were excited at the equator and propagated to higher latitudes, even though the wavenormal angles at the equator were estimated via ray-tracing to be close to the resonance cone. At the point of observation, however, the 26 degree latitude event had a wavenormal angle 15-20 degrees inside the local resonance cone.

There is a local time difference in the spectral densities for the nominally electrostatic wave band, $(1-3)f_{ce}$, and the highest whistler band, $(0.5-1.0)f_{ce}$. The upper whistler band has the highest mean spectral densities in the 3-6 LT sector, for L=5-7. The mean spectral densities per visit to these regions are peaked in the 10^{-9} to 10^{-11} v^2/m^2Hz range. In contrast, the largest mean values for the $(1-3)f_{ce}$ band are more extended in LT, particularly in the midnight sector, from 21 through 09 LT (Figure 11 (ECH) and Figure 5 (whistler)). Roeder and Koons (1989) expected a local time difference between the occurrence of ECH waves and whistler waves. Since the two types of waves have different energy requirements for amplifying particles, the LT pattern would be depend on the convection patterns of the two particle populations.

ECH Waves

A great deal of work has been done with regard to the narrow-band nature of ECH waves, their tendency to intensify at the equator, their amplitudes and their capacity to precipitate electrons and cause the diffuse aurora (Shaw & Gurnett, 1975; Christiansen et al., 1978; Hubbard & Birmingham, 1978 and Roeder & Koons, 1989). More recently, CRRES data were used to study the intensities of the harmonics (Paranicas et al., 1992). Belmont et al. (1983) and Roeder & Koons (1989) concluded from satellite surveys (GEOS 2, SCATHA and AMPTE IRM) that sufficiently large amplitudes rarely occurred, and ECH waves could not be considered a significant contributor to the diffuse aurora. The spatial distributions of power in the $(1-3)f_{ce}$ band in this survey are in general agreement with the ECH survey of Roeder & Koons (1989). CRRES spent more time inside of geosynchronous orbit than either SCATHA or IRM, improving sampling there at latitudes up to 28 degrees.

Filbert & Kellogg (1988) used the work of Ashour-Abdalla et al. (1979), together with their calculation of ECH wavenumbers, to estimate the energy of the warm electrons responsible for ECH wave amplification to be between tens and hundreds of eV. Belmont et al. (1983) calculated the minimum wave amplitude for strong difffusion at L=7. For the cases of resonant particle energies of 200 eV and 1 keV, they calculate amplitudes of 0.6 mV/m and 2 mV/m, respectively. Roeder & Koons (1989) estimate from the calculations of Belmont et al. (1983) and Coroniti (1985) that ECH amplitudes for strong diffusion for 1 keV electrons in the L=4-8 range should be 1-14 mV/m (the higher is for lower L values). In none of the equatorial regions sampled by CRRES was the mean amplitude per visit greater than 0.6 mV/m.

The $(1-3)f_{ce}$ band measurements include the contributions from broadband features as well as those from the ECH emissions. Some examples of the broadband waves may be found

in orbit 0515, Figure 1, from 1200 to 1240 UT. These impulsive waves extended from below f_{ce} up to approximately the continuum radiation. They were observed at all latitudes. Nevertheless, the results of this survey are consistent with the conclusions of Belmont et al. (1983) and Roeder & Koons (1989) that amplitudes in this band are rarely strong enough to cause the strong diffusion expected for the diffuse aurora.

Roeder & Koons (1989) also noted a latitudinally isolated population of ECH waves at 18-20 degrees latitude. CRRES observed ECH enhancements up to $10^{-9.7}$ v^2/m^2Hz at latitudes up to 25 degrees. In the midnight sector, these events helped raise the off-equator mean spectral densities to within two orders of magnitude of the largest mean values on the equator.

Summary

In the two highest whistler bands, $(0.3-0.5)f_{ce}$ and $(0.5-1.0)f_{ce}$, the highest mean spectral densities were in the 03-09 LT sector for L=5-7. The power in the $(0.5-1.0)f_{ce}$ band is concentrated within about 10 degrees of the equator. The local time pattern is consistent with amplification by substorm-injected electrons.

The lowest whistler band also shows a dawnside peak in mean spectral density, however the peak moves around to the dayside with increasing latitude. For L=4-7, the mean spectral densities in the 15-29 degree dayside region exceed those within five degrees of the equator. These waves may be escaping from the plasmasphere, or coming from the minima in magnetic field at high latitudes on the dayside, or from a distant equatorial source beyond the plasmasphere. Since the mean intensity falls off for L=7-8 on the dayside, the minimum-B regions may not be the dominant source.

In midnight-dawn local times, there is a roughly proportional relation between the mean spectral density per visit and Kp at the time of visit. This relation vanishes for 12-21 LT. If it takes more than three hours for injected electrons to drift to noon LT and amplify whistler waves, the Kp at the time of amplification would not be the same as Kp at the time of injection. This drift time is consistent with electron energies of tens of keV at L=5-6.

In the ECH band above f_{ce}, the largest mean intensity is in the midnight-dawn sector, within 5 degrees of the equator. The largest mean amplitude per visit was on the equator in the midnight sector and L=6-7, at 0.6 mV/m. The 15-18 sector had the weakest ECH waves. In some cases, they were not detectable at the 10^{-17} v^2/m^2Hz level in this region. Occasional off-equator intensifications of ECH waves contributed to off-equator midnight mean spectral densities that were only two orders of magnitude less than the largest equatorial mean values.

References

Anderson, R.R. and D.A. Gurnett, "The Development of Static and Dynamic Models od the Earth's Radiation Belt Environmet Through the Study of Plasma Waves, Wave-Particle Interactions and Plasma Number Densities from In Situ Observations in the Earth's Magnetosphere with the CRRES Spacerad Instruments", Scientific Report PL-TR-92-2035, Phillips Laboratory, 1992. ADA251245

Belmont, G., D. Fontaine and P. Canu, "Are Equatorial Electron Cyclotron Waves Responsible for Diffuse Auroral Electron Precipitation?", Journal of Geophysical Research, 88, 9163-9170, 1983.

Burtis, W.J. and R.A. Helliwell, "Magnetospheric Chorus: Occurrence Patterns and Normalized Frequency", Planetary and Space Science, 24, 1007-1024, 1976.

Christiansen, P.J., M.P. Gough, G. Martelli, J.J. Bloch, N. Cornilleau, J. Etcheto, R. Gendrin, C. Beghin, P. Decreau and D. Jones, "GEOS-1 Observations of Electrostatic Waves, and Their Relationship with Plasma Parameters", Space Science Reviews, 22, 383, 1978.

Davidson, G.T., "The Motion of Charged Particles in the Earth's Magnetic Field", in *The Trapped Radiation Handbook*, ed. J.B. Cladis, G.T. Davidson and L.L. Newkirk, Lockheed Palo Alto Research Laboratory, DNA 25241H, 1977.

Dunckel, N. and R.A. Helliwell, "Whistler Mode Emissions on the OGO 1 Satellite", Journal of Geophysical Research, 74, 6371-6385, 1969.

Dunckel, N. and R.A. Helliwell, "Whistler Mode Emissions on the OGO 1 Satellite", Journal of Geophysical Research, 74, 6371-6385, 1969.

Filbert, P.C. and P.J. Kellogg, "On the Wavelengths of $(n+1/2)$fce Gyroharmonic Emissions in the Earth's Magnetosphere", Journal of Geophysical Research, 93, 11374-11384, 1988.

Hayakawa, M., Y. Yamanaka, M. Parrot and F. LeFeuvre, "The Wave Normals of Magnetospheric Chorus Emissions Observed on Board GEOS 2", Journal of Geophysical Research, 89, 2811-2821, 1984.

Heppner, J.P., M.C. Liebrecht, N.C. Maynard and R.F. Pfaff, "High Latitude Distributions of Plasma Waves and Spatial Irregularities from DE-2 AC Electric Field Observations", submitted to ....,1992.

Huang, C.Y., C.K. Goertz and R.R. Anderson, "A Theoretical Study of Plasmaspheric Hiss Generation", Journal of Geophysical Research, 88, 7927-7940, 1983.

Inan, U.S., Y.T. Chiu and G.T. Davidson, "Whistler-Mode Chorus and Morningside Aurorae", Geophysical Research Letters, 19, 653-656, 1992.

Koons, H.C. and J.L. Roeder, "A Survey of Equatorial Magnetospheric Wave Activity Between 5 and 8 R_E", Planetary and Space Science, 38, 1335-1341, 1990.

Lyons, L.R. and R.M. Thorne, "The Magnetospheric Reflection of Whistlers", Planetary and Space Science, 18, 1753-1767, 1970.

Lyons, L.R., R.M. Thorne and C.F. Kennel, "Pitch Angle Diffusion of Radiation Belt Electrons Within the Plasmasphere", Journal of Geophysical Research, 77, 3455, 1972.

Maeda, K., "Cyclotron Side-Band Emissions from Ring Current Electrons", Planetary and Space Science, 24, 341-347, 1976.

Muto, M. H. Hayakawa, M. Parrot and F. LeFeuvre, "Direction-Finding of Half-Gyrofrequency VLF Emissions in the Off-Equatorial Region of the Magnetosphere and Their Generation and Propagation", Journal of Geophysical Research, 92, 7538-7550, 1987.

Muzzio, J.L. and J.J. Angerami, "OGO 4 Observations of Extremely Low Frequency Hiss", Journal of Geophysical Research, 77, 1157-1173, 1972.

Paranicas, C., W.J. Hughes, H.J. Singer and R.R. Anderson, "Banded Electrostatic Emissions Observed by the CRRES Plasma Wave Experiment", Journal of Geophysical Research, 97, 13889-13898, 1992.

Parrot, M. and F. LeFeuvre, "Statistical Study of the Propagation Characteristics of ELF Hiss Observed on GEOS-1, Inside and Outside the Plasmasphere", Annales Geophysicae, 4, 363-384, 1986.

Roeder, J.L. and H.C. Koons, " A Survey of Electron Cyclotron Waves in the Magnetosphere and the Diffuse Auroral Electron Precipitation", Journal of Geophysical Research, 94, 2529, 1989.

Russell, C.T. and R.E. Holzer, "OGO 3 Observations of ELF Noise in the Magnetosphere 1. Spatial Extent and Frequency of Occurrence", Journal of Geophysical Research, 74, 755-777, 1969.

Shaw, R.R. and D.A. Gurnett, "Electrostatic Noise Bands Associated with the Electron Gyrofrequency and Plasma Frequency in the Outer Magnetosphere", Journal of Geophysical Research, 80, 4259, 1975.

Singer, H.J., W.P. Sullivan, P. Anderson, F. Mozer, P. Harvey, J. Wygant and W. McNeil, "Fluxgate Magnetometer Instrument on the CRRES", Journal of Spacecraft and Rockets, 29, 599-601, 1992.

Sonwalkar, V.S. and U.S. Inan, "Wave Normal Direction and Spectral Properties of Whistler Mode Hiss Observed on the DE-1 Satellite", Journal of Geophysical Research, 93, 7493-7514, 1988.

Thorne, R.M., E.J. Smith, R.K. Burton and R.E. Holzer, "Plasmaspheric Hiss", Journal of Geophysical Research, 78, 1581-1596, 1973.

Tsurutani, B.T. and E.J. Smith, "Postmidnight Chorus: A Substorm Phenomenon", Journal of Geophysical Research, 79, 118-127, 1974.

Tsurutani, B.T. and E.J. Smith, "Two Types of Magnetospheric ELF Chorus and Their Substorm Dependences", Journal of Geophysical Research, 82, 5112-5128, 1977.
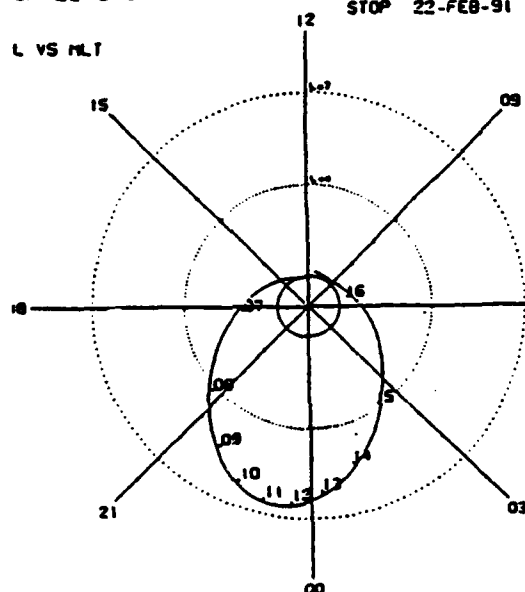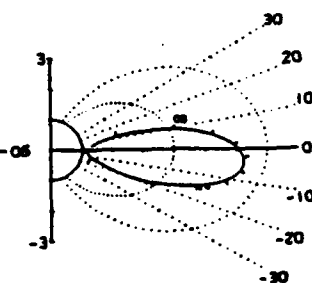
FIGURE 1

394

FIGURE 2

ORBIT 515   2/22 1991   09:09:59.3400

FIGURE 3

FIGURE 4

FIGURE 6

398

FIGURE 5

4-OCT-91 0512 Orbit 1049 CRRES Iowa Plasma Waves
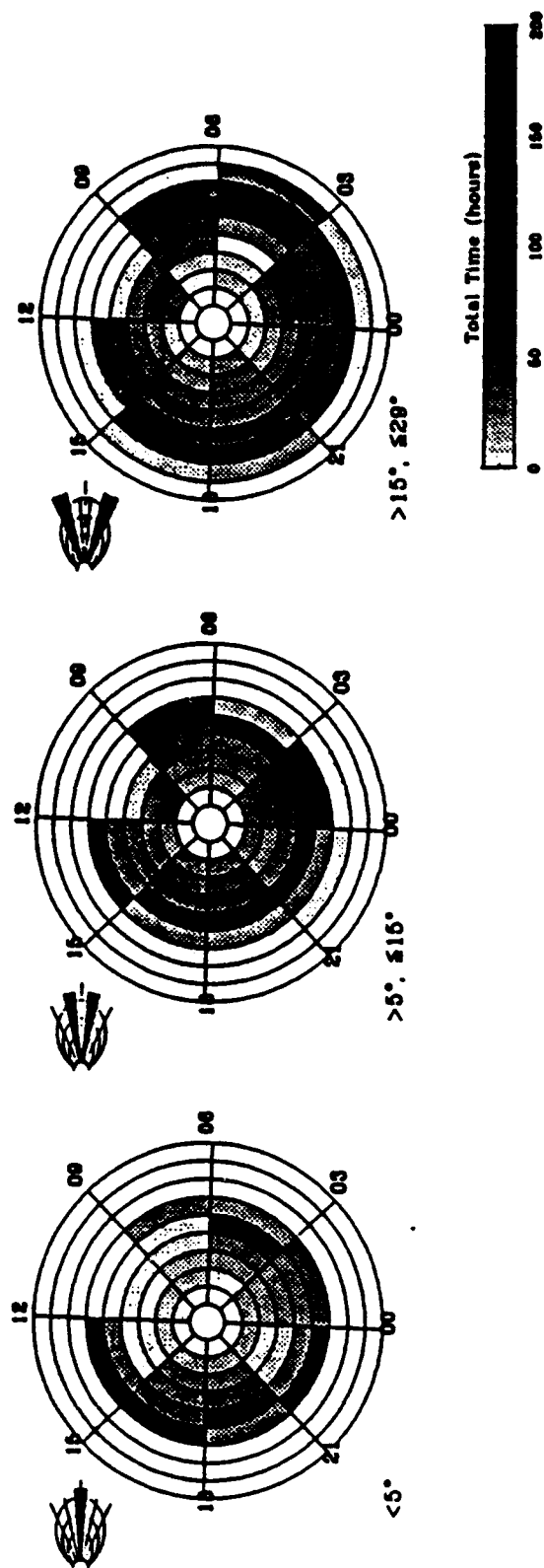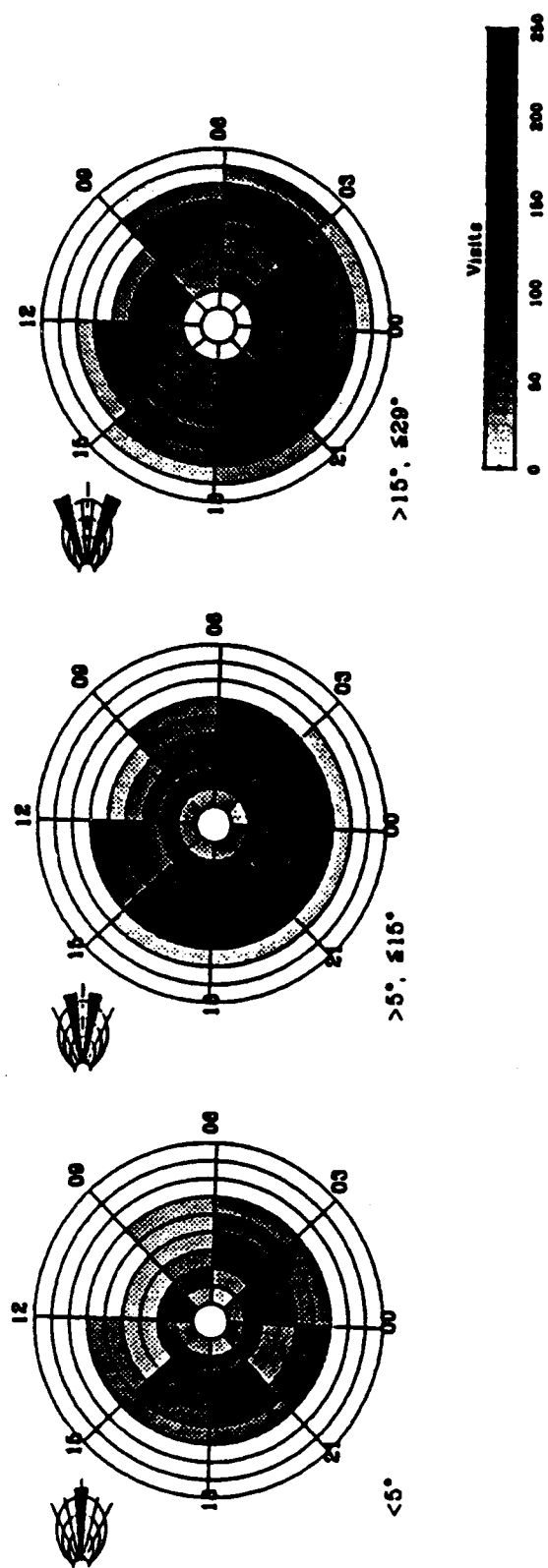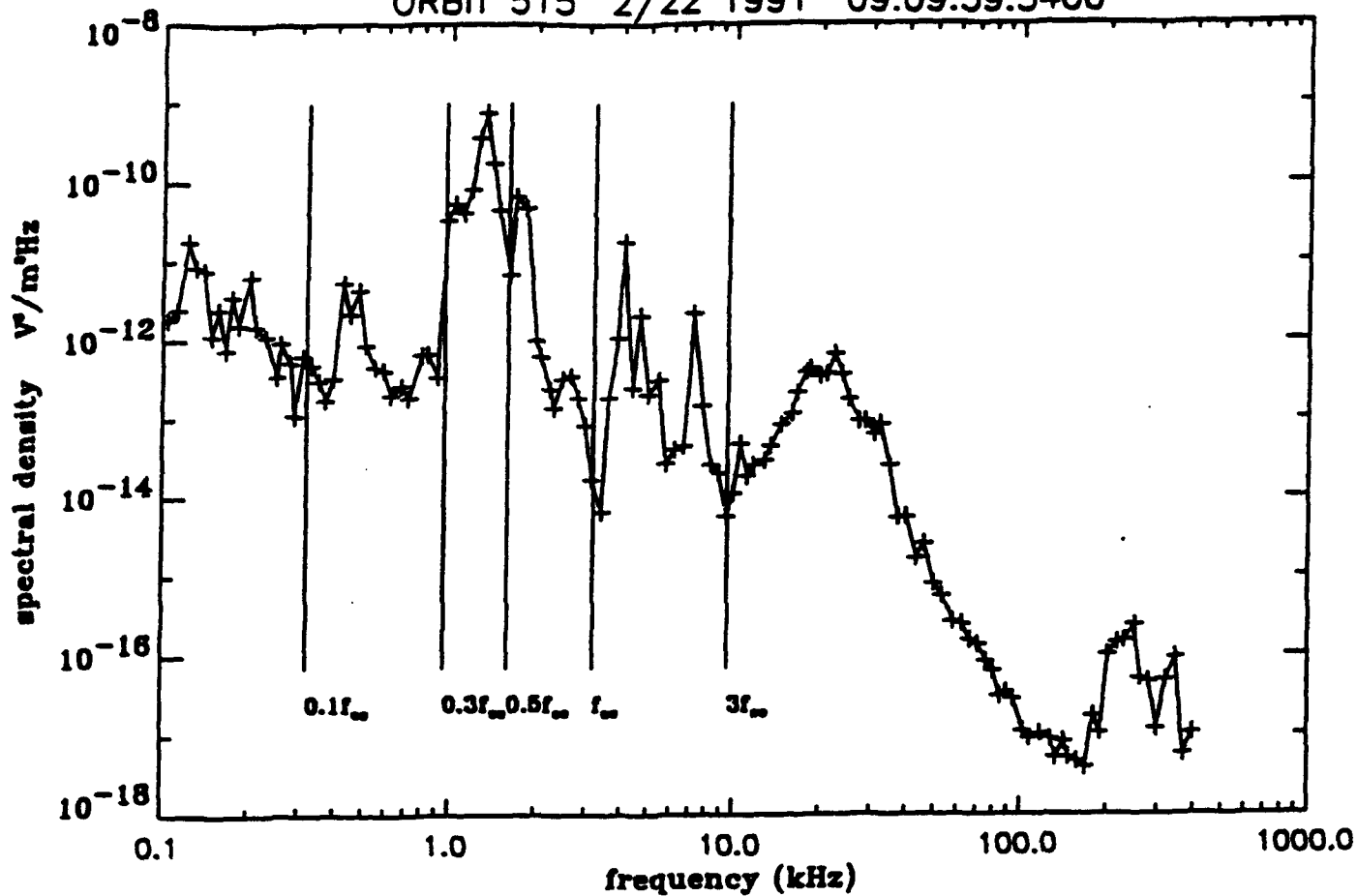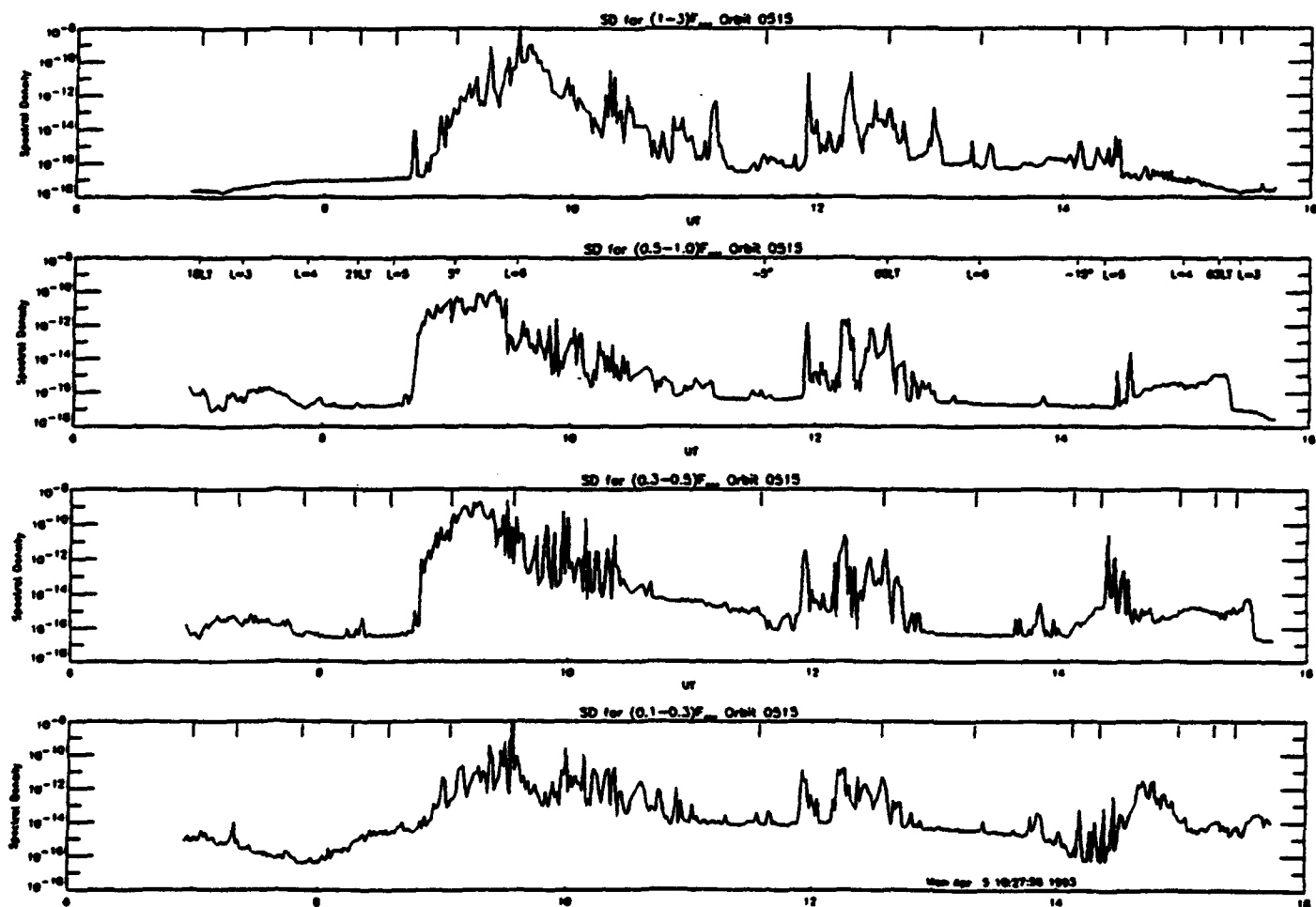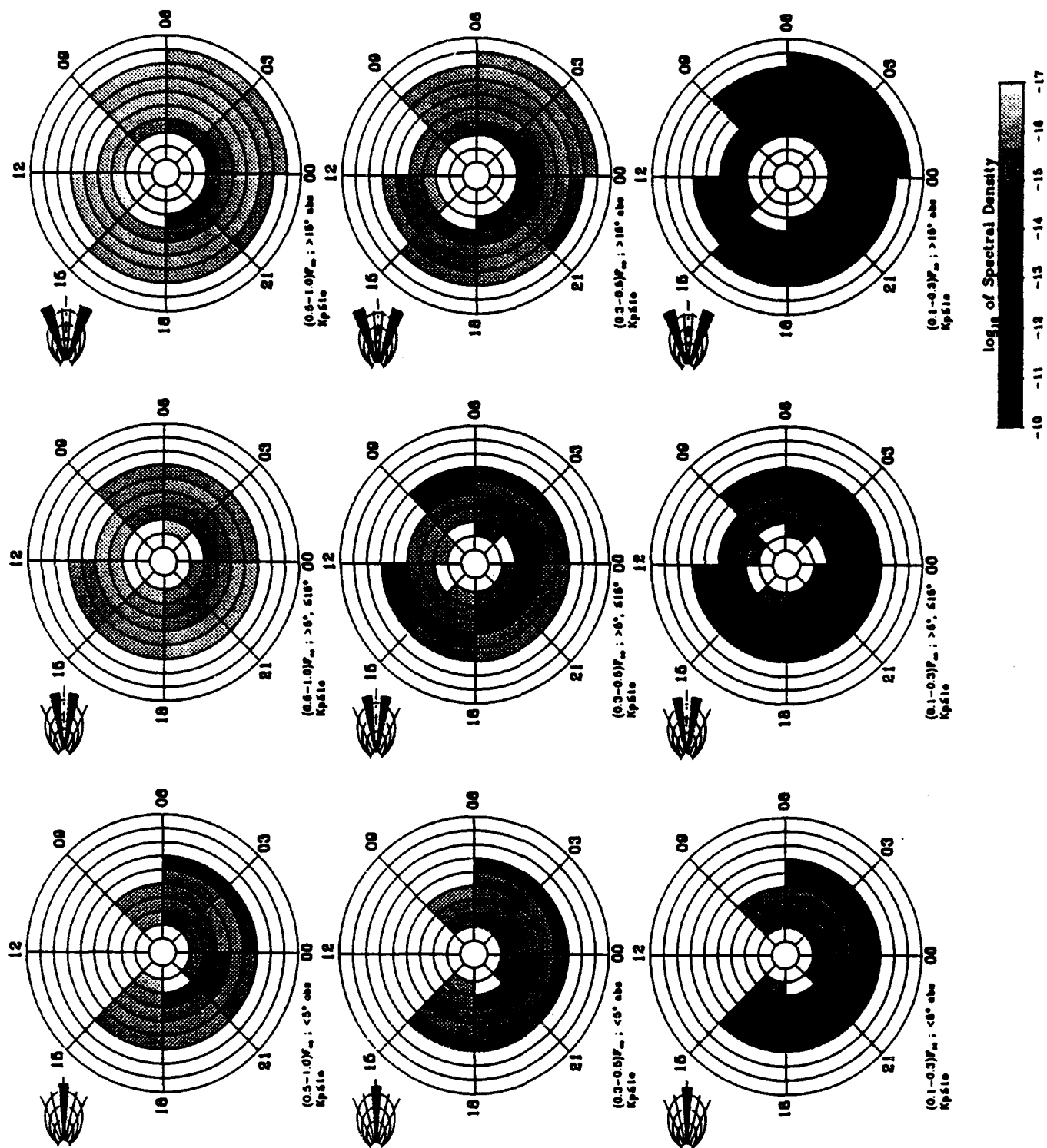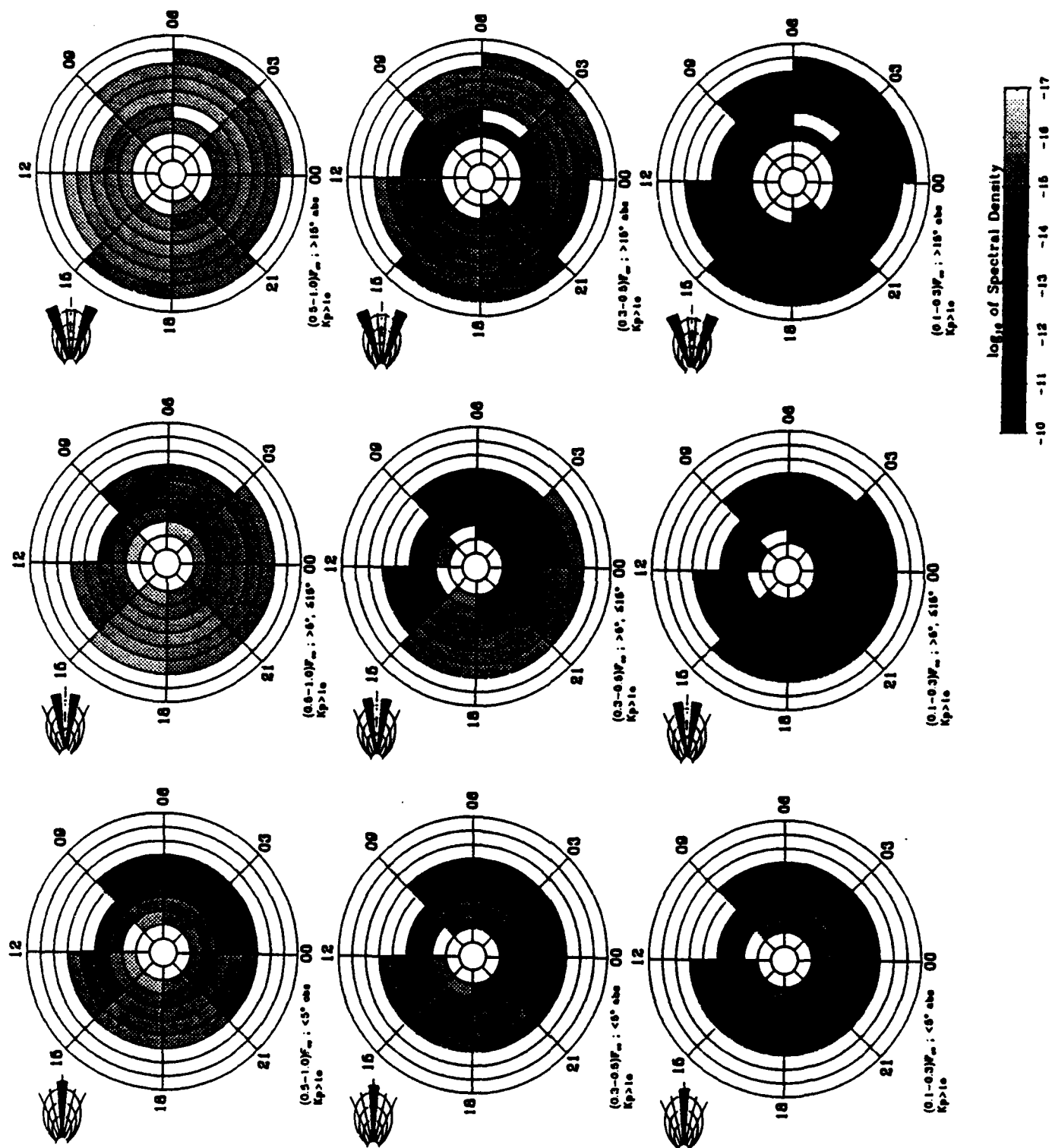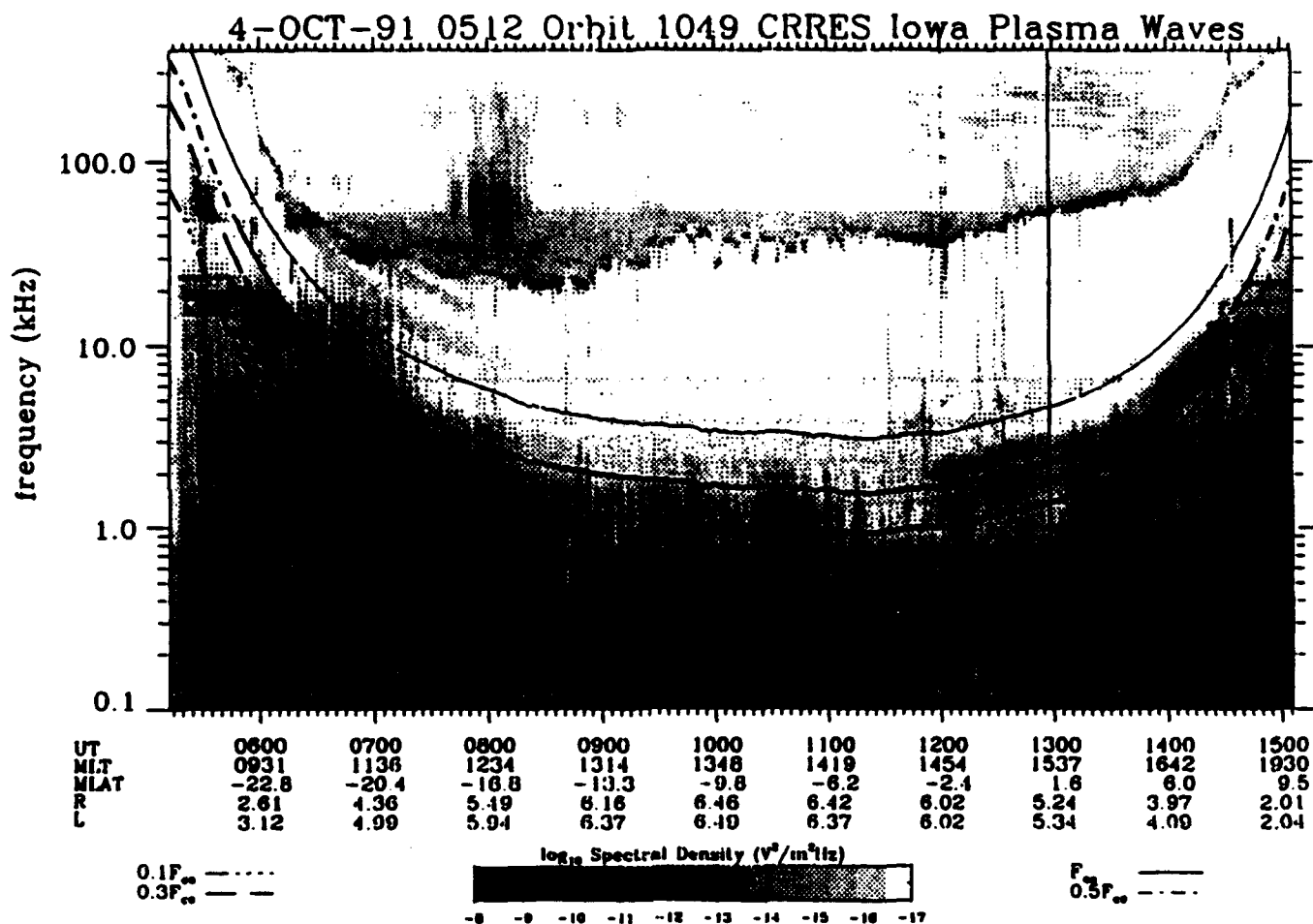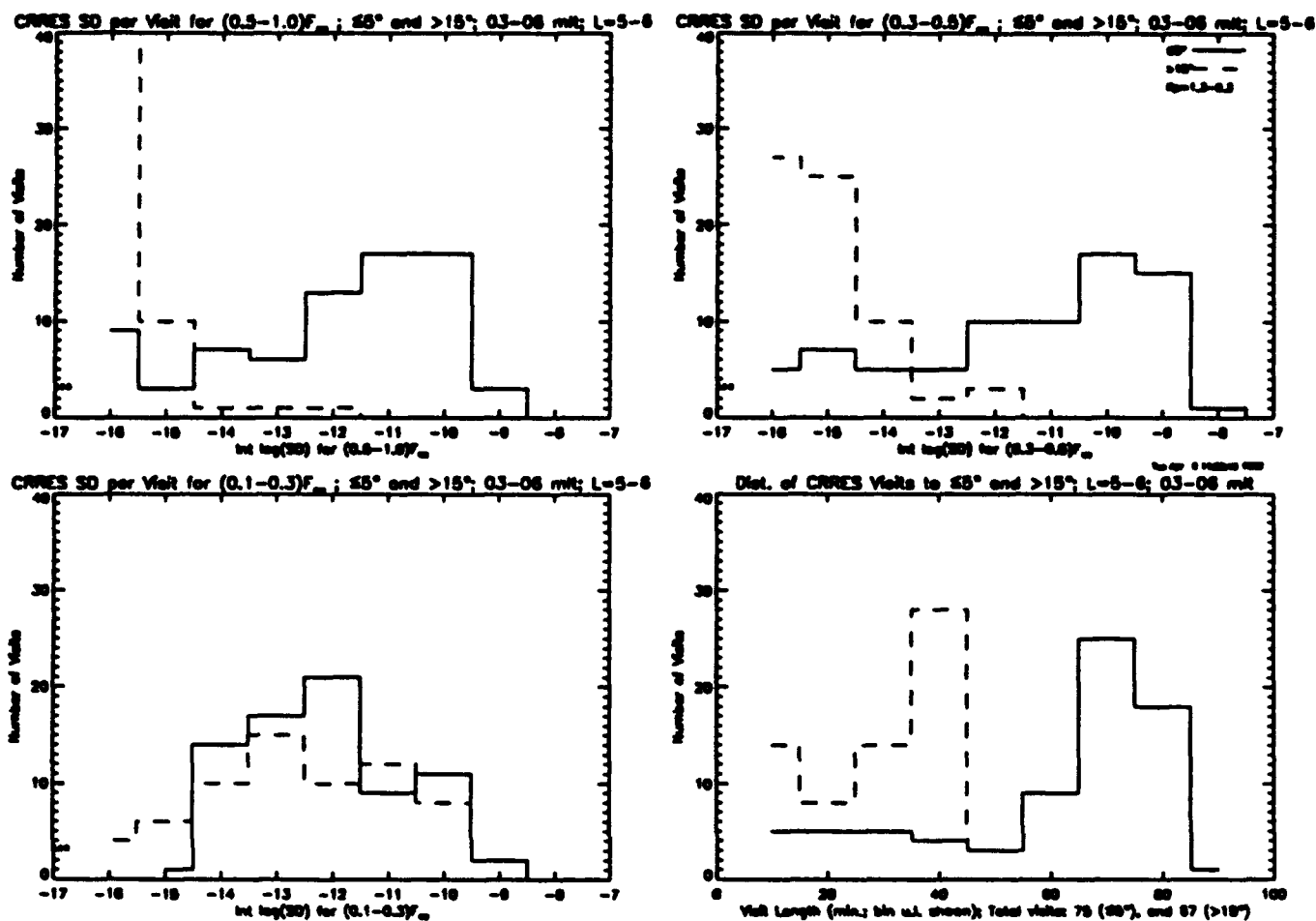
FIGURE 7

FIGURE 8

FIGURE 9

FIGURE 10

402

FIGURE 11

403

CRRES SD per Visit for $(1-3)F_m$; ≤5° and >15°; 00—03 mlt; L=6—7

Dist. of CRRES Visits to ≤5° and >15°; L=6—7; 00—03 mlt

FIGURE 12

12-APR-91 1025 Orbit 0635 CRRES Iowa Plasma Waves
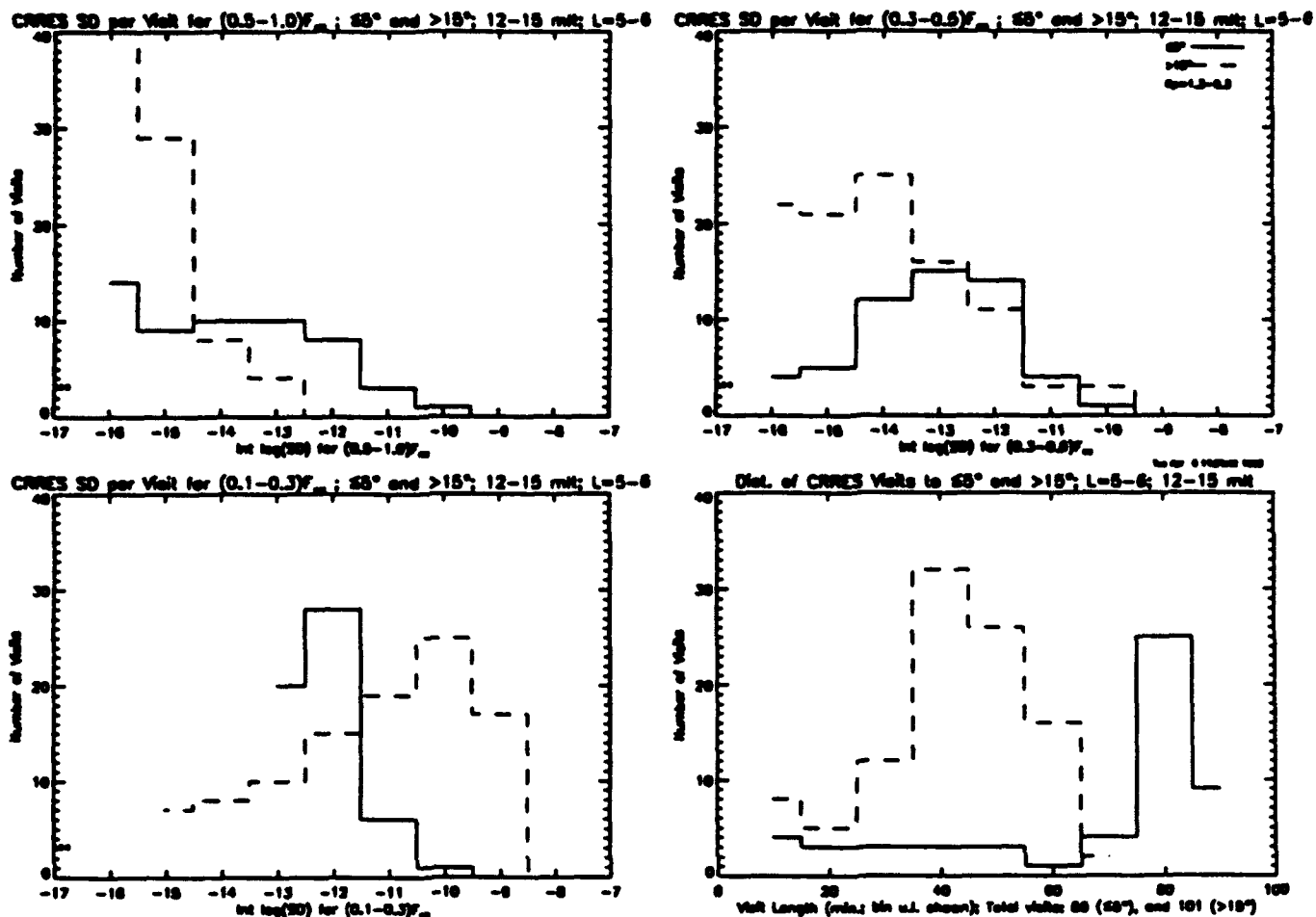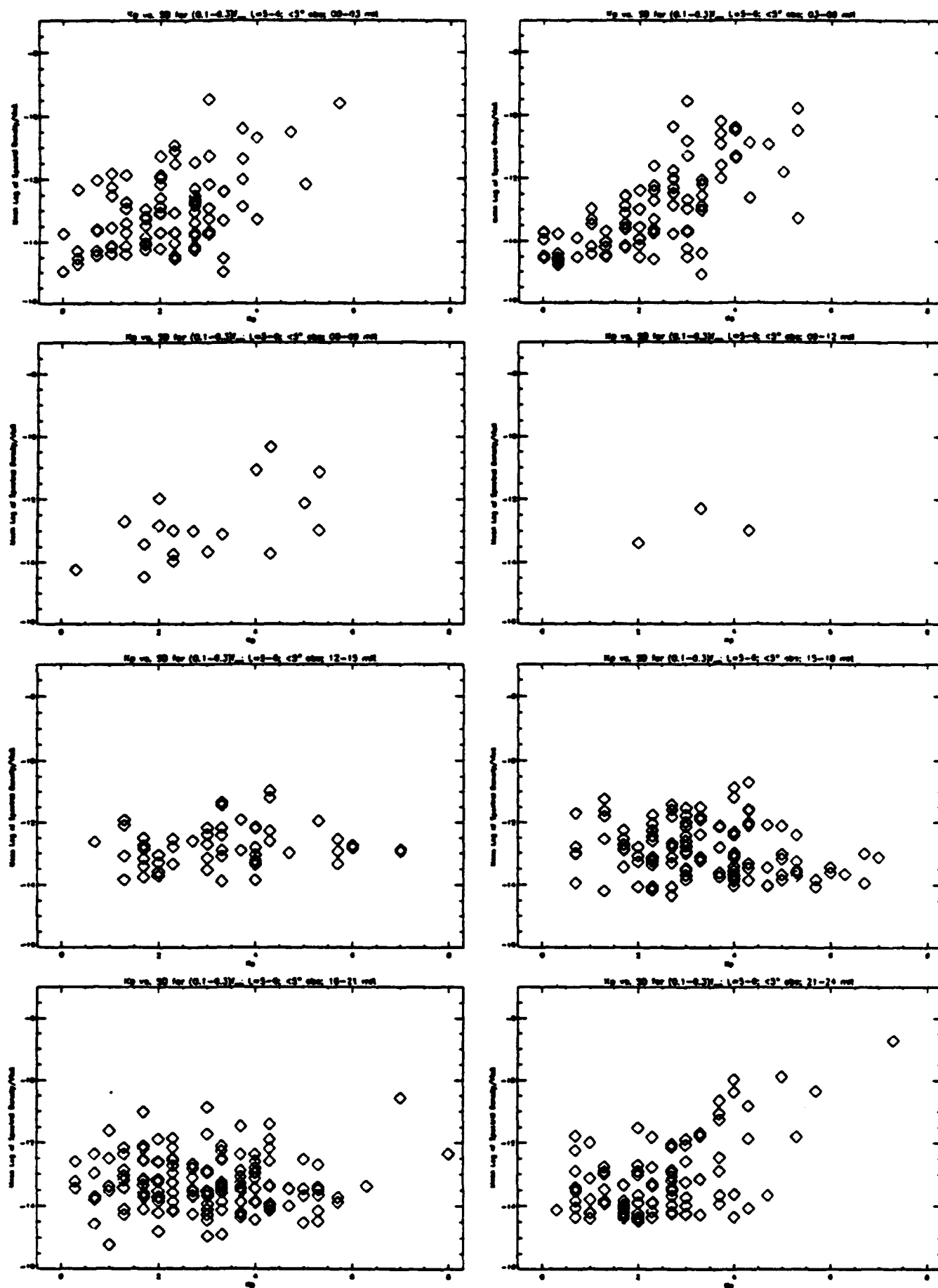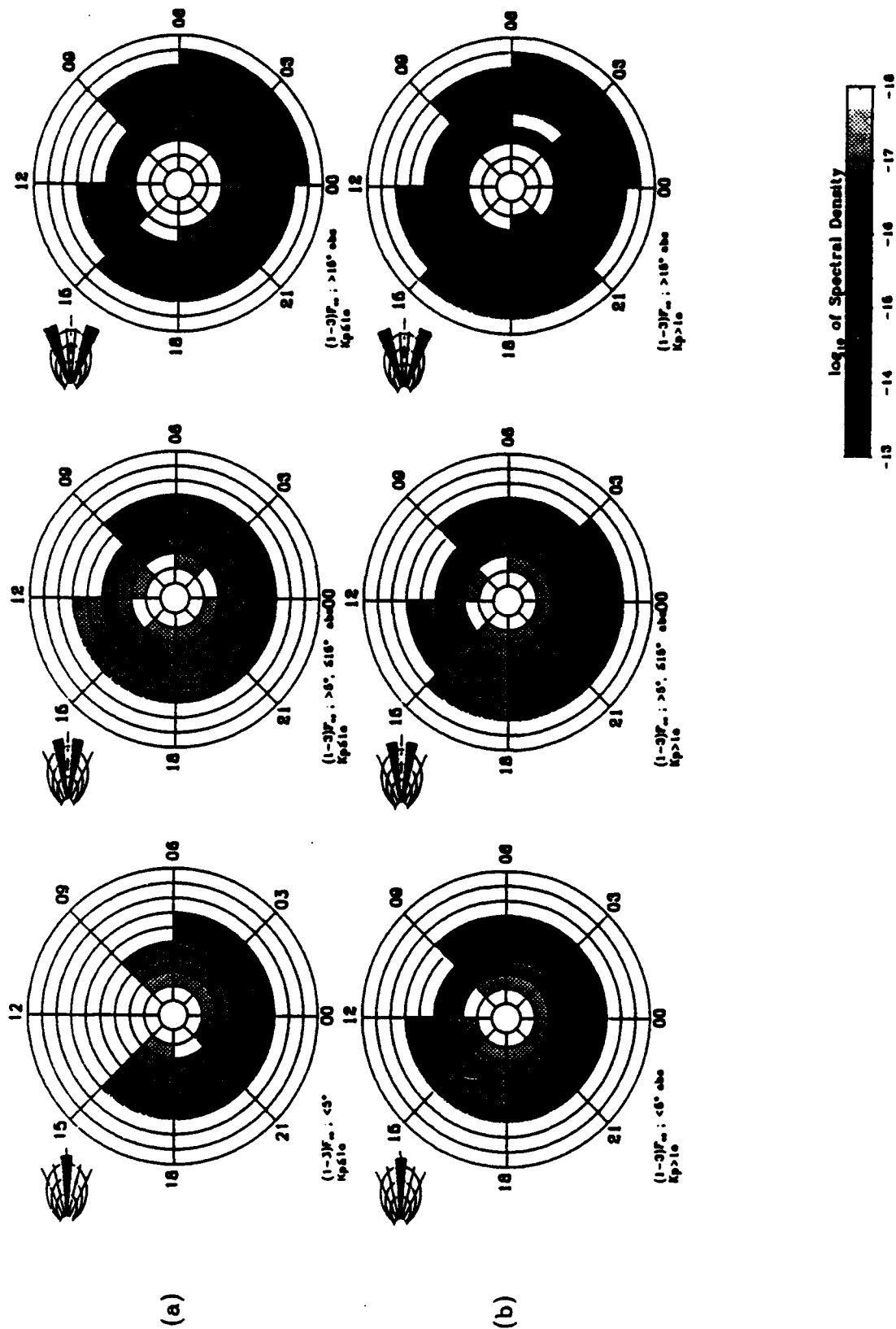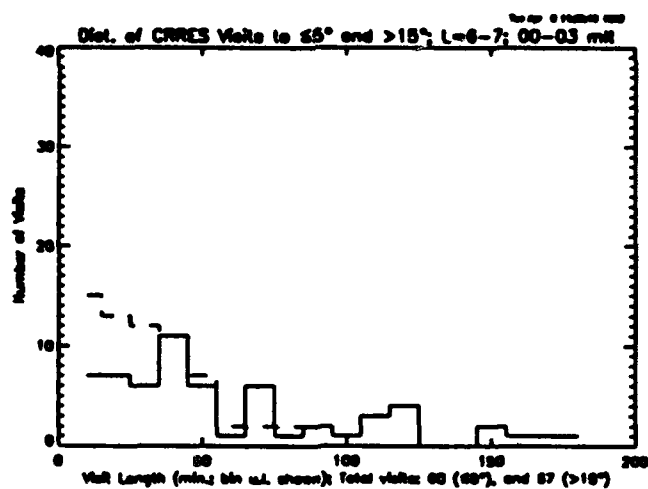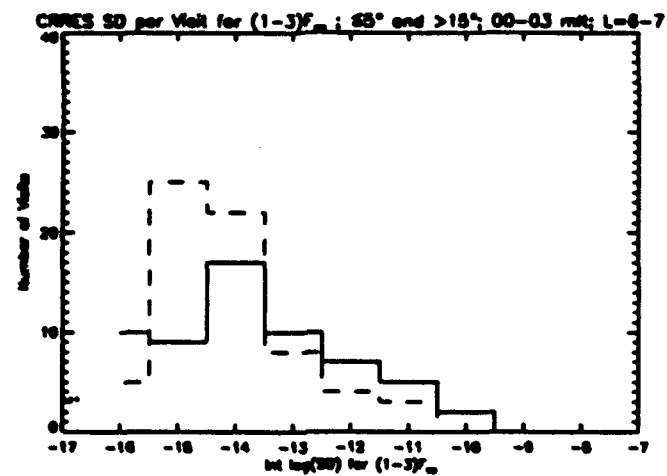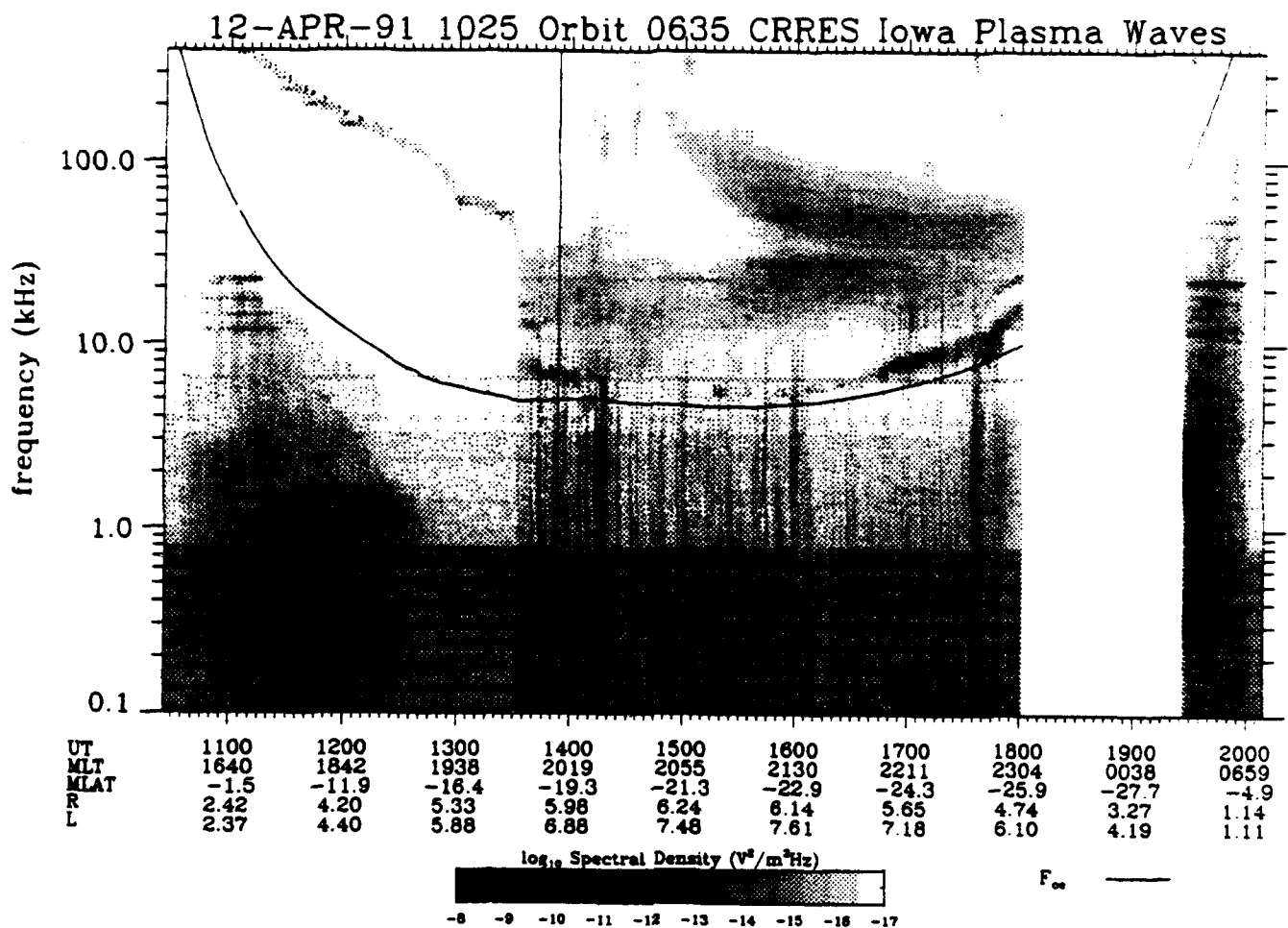
FIGURE 13

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

STUDY OF INFRARED EMISSIONS FROM H II REGIONS

Prepared by: Dr. Gregory C. Sloan

Research Location: Phillips Laboratory
                                      Geophysics Directorate
                                      Hanscom AFB, MA 01731

AF Research Colleague: Frank O. Clark and
                                         Stephan D. Price

Contract Number: F19628-86-C-0224

Final Report

for Southeastern Center for
Electrical Engineering Education

Gregory C. Sloan

Contract: F19628-86-C-0224
Period: 1 April 1992 — 30 June 1993

1. Introduction

From 1 April, 1992 to 30 June, 1993 I was employed by the Southeastern Center for Electrical Engineering Education (SCEEE) as a Geophysics Scholar for the Geophysics Directorate of Phillips Laboratory, Optical Environment Division, Backgrounds Branch. My effort focal points were Stephan D. Price and Frank O. Clark.

During this period I was involved in a variety of projects, as described below. In all, I attended three conferences and published four papers. Additional publications based on work done during the last year should be forthcoming.

2. Effort Activities

When I arrived at the Geophysics Directorate, I first prepared two chapters of my dissertation for publication. While working on these, two previously submitted papers appeared in press (LeVan et al. 1992; Landau, Grasdalen & Sloan 1992). My dissertation focused on observations made with the Air Force Geophysics Lab. Infrared Array Spectrometer (GLADYS), reduction methods developed for data from this instrument, and results for some celestial objects. The two chapters submitted in the summer of 1992 covered the results of maximum entropy reconstructions of images of α Orionis, a late-type oxygen-rich supergiant with an extended dust shell (Sloan, Grasdalen, and LeVan 1993a) and the Red Rectangle, a bipolar nebula centered on the star HD 44179, well known as a source of the unidentified infrared (UIR) emission features (Sloan, Grasdalen and LeVan 1993b). An additional chapter, describing in detail our maximum entropy reconstruction algorithm is still in preparation (Sloan and Grasdalen 1993).

While these articles were being prepared, I also attempted maximum entropy reconstruction of GLADYS data taken in February, 1991 of the planetary nebula NGC 7027. This nebula is another well-known source of the UIR emission features and is spatially extended on a scale of several arcseconds. Unfortunately, the signal/noise of the data was not high, partly because they were taken with the slit oriented north/south, instead of NE/SW along the brightest regions of the nebula. As a result, the reconstructions did not provide sufficient spatial resolution to justify publishing the results.

I devoted most of the summer to in-house projects for Steve Price and Frank Clark. I first prepared a catalog of known sources of UIR emission (roughly 100), and then began to work on a more extensive catalog of bright infrared sources. Concurrent with this was an effort to check the spectral templates of Martin Cohen for five standard stars. These templates were constructed by splicing spectra from a variety of sources together to cover the wavelength range 2-30 μm. They provide the best-calibrated spectra available for the standard stars α Lyrae, α Canis Majoris, α

Boötes, α Tauri, and β Pegasi. By combining photometric data from the Catalog of Infrared Observations (CIO; Gezari, Schmitz, and Mead 1987) and the spectral database from the Infrared Astronomical Satellite Low-Resolution Spectrometer (IRAS LRS), I was able to show that Cohen's spectral templates were at least as good as the available data. These comparisons were presented at the Infrared Celestial Backgrounds Review held at Utah State University 14-15 July, 1992. Work on the catalog of bright infrared sources eventually culminated in a catalog of 348 sources with 12 μm fluxes brighter than a magnitude of -2. We are still planning to use the CIO and LRS to produce spectral energy distributions for all of the sources so that we may develop classification schemes from the database.

One small project pursued at this time was assisting with a proposal by Gary Grasdalen, J.A. Benson, and Tom Hayward to refurbish the 10 and 20 μm 6-channel spectrometers at Wyoming by installing two-dimensional arrays. This proposal did not receive funding from the National Science Foundation. A more involved effort was a proposal with Charles Woodward to observe the Red Rectangle with a long-slit spectrometer at 3.3 and 3.4 μm at Kitt Peak National Observatory. The initial proposal to observe in the spring of 1993 was rejected, but we resubmitted the proposal and received two nights for November, 1993.

The departure of Paul LeVan from the Geophysics Directorate made it necessary for me to step in and take his place in the project to refurbish GLADYS. The plan was to completely replace all of the electronics in this instrument, the clock-drives, the array inputs, the pre-amps, the analog-digital converters, and the co-adder electronics, with equipment from Wallace Instruments in Ithaca, New York. This project also involved Peter Tandy, an engineer here at Phillips, and Bruce Pirger, a summer student from Cornell. By the end of the summer the electronics were working on their own, and in the fall we integrated the electronics with the detector array. Gary Grasdalen and James Weger flew out from Wyoming in November to assist in a series of tests to calibrate the system and determine its capabilities. Initial estimates that we had solved serious problems with the older system and improved the noise by a factor of ten later proved to be correct.

While working with GLADYS, I also began a study of an interesting 13 μm emission feature associated with the strong silicate emission feature at 10 μm seen in some of the LRS spectra. This investigation has continued through the winter and is now approaching publication. Preliminary results were presented at the summer meeting of the American Astronomical Society (AAS) in Berkeley (LeVan, Sloan, & Little-Marenin, 1993). We have found 73 sources of the 13 μm emission feature. The typical source is a giant of spectral class M6 or M7, a semi-regular variable (SRb), and has an LRS characterization of 14-15 or 21-24 (weak silicate emission superimposed on a stellar continuum). When we complete work on comparison samples of galactic Mira and SRb variables, we will be ready to publish our results.

For the most part, my work since the fall has focused on GLADYS. Our first observing run at the Wyoming Infrared Observatory (WIRO) was scheduled for late February, and a tremendous amount of lab work and programming had to be completed before then. One break was the winter AAS meeting in Phoenix, where I presented my methods of data reduction for long-slit spectroscopy and the results from my dissertation (Sloan 1993). A second diversion was the preparation of a proposal to the NASA Astronomical Data Program with Bob Stencel and Martin Cohen to study the LRS database to better understand the nature of stellar photospheres and dust shells in the 7-25 μm regime. Finally, during this period I also prepared a *Report of Operations of the Air Force Geophysics Laboratory Infrared Array Spectrometer*, presently in press as technical report PL-TR-93-2012 (Sloan, LeVan, and Tandy, 1993)

The first observing run at WIRO was a limited success. Peter Tandy and I were met by Bruce Pirger in Laramie during the week of 20-26 February. Pirger had brought with him the last of the hardware to integrate with GLADYS and by the first night of the run, the integration was complete. The run, which ran from 27 February to 5 March, was flawed by poor observing conditins and by technical problems typical for a first run with new equipment. We did manage to obtain data on a variety of objects. We confirmed that the noise equivalent flux density was about 10 Jy Hz$^{-1/2}$, as estimated in the fall. This meant that we had improved the system noise by a factor of 10 over the old electronics.

Preliminary reductions of the data from this run were presented at the summer AAS conference in Berkeley (Sloan et al. 1993). These consisted of both spectra and spatiograms (plots of the width of an object as a function of wavelength) for AFGL 2688 and IRC+10216, both highly evolved carbon-rich stars, and α Orionis. Further processing of the data await the solution of our calibration difficulties. The flatfielding algorithm is complicated by the non-linear nature of the response functions of the pixels on the array. A rough algorithm has been developed, but a more precise method requires dark current measurements. Since GLADYS has no dark slide, this entails opening and temporarily modifying the dewar. Once the data can be properly calibrated, we will be able to apply maximum entropy reconstruction to these spatially extended sources. This step should produce some exciting results.

Peter Tandy and I returned to Wyoming, working in Laramie 21-25 June, and observing at WIRO 26-30 June. This second run was flawless from an engineering standpoint, but again the weather was marginal. Most of the data obtained consisted of careful observations of standard stars to test the assertion of Martin Cohen and collaborators that the spectra of late-type stars have an absorption feature at 8 μm due to SiO absorption. This is evolving into a separate project to study the dependence of this absorption with spectral type and to parameterize it with physical models. We obtained additional data on NGC 7027 (with a NE/SW slit), and other planetary nebulae such as IRAS 21282+5050. We are planning to apply maximum

entropy reconstruction to these data in the future. Further observations, made in collaboration with Charles Woodward, were of Nova Aquila 1993. We plan to monitor this and other novae in the future.

An analysis of the signal/noise aspects from this run indicates that the noise was higher than it was in March. We suspect that the co-adder may be noisier when operating at higher temperatures. It has become clear that the system is not background limited. The noise equivalent flux density is roughly a factor of 4 larger than that expected from counting noise from the flux from the telescope and sky. Hopefully, we will be able to identify and eliminate some of the noise sources in the co-adder in the near future.

3. Conclusion

The Geophysics Scholar program has been replaced by the Phillips Laboratory Scholar program, administered by the Northeast Consortium for Engineering Education. I am now supported by this new program and will continue the projects I have described. Some of these are nearing fruition and their results will soon be prepared for publication. First among these is the study of sources of 13 μm emission. The calibration issues with GLADYS should soon be understood, and this will allow completion of the studies of silicon monoxide in late-type giants, and the spatial variations in spectral behavior of the evolved stars AFGL 2688 and IRC+10216 and the planetary nebulae NGC 7027 and IRAS 21282+5050.

References

Gezari, D.Y., Schmitz, M., & Mead, J.M. 1987, *Catalog of Infrared Observations*, NASA Reference Publication 1196.
Landau, R., Grasdalen, G.L., & Sloan, G.C. 1992, *A&A*, **259**, 696.
LeVan, P.D., Sloan, G.C., & Little-Marenin, I.R. 1993, *BAAS*, **25**, 877.
LeVan, P.D., Sloan, G.C., Little-Marenin, I.R., Grasdalen, G.L., 1992, *ApJ*, **392**, 702.
Sloan, G.C. 1992, *BAAS*, **24**, 1302.
Sloan, G.C., & Grasdalen, G.L. 1993, in preparation.
Sloan, G.C., Grasdalen, G.L., & LeVan, P.D. 1993a, *ApJ*, **404**, 328.
_____., 1993b, *ApJ*, **409**, 412.
Sloan, G.C., LeVan, P.D., and Tandy, P.C. 1993, *Report of Operations of the Air Force Geophysics Laboratory Infrared Array Spectrometer*, PL-TR-93-2012.
Sloan, G.C., Tandy, P.C., Pirger, B.E., & Hodge, T.M. 1993, *BAAS*, **25**, 876.

Appendices

Papers published 1 April 1992 — 30 June 1993

LeVan, P.D., Sloan, G.C., Little-Marenin, I.R., Grasdalen, G.L., 1992, "8-14 Micron Spectroscopy of Carbon Stars Associated with Silicate Dust," *ApJ*, **392**, 702.

Landau, R., Grasdalen, G.L., & Sloan, G.C. 1992, "Three-beam Chopping: an Efficient Infrared Observing Technique," *A&A*, **259**, 696.

Sloan, G.C., Grasdalen, G.L., & LeVan, P.D. 1993a, "Spatially Resolved Spectra of Silicate Dust around α Orionis," *ApJ*, **404**, 328.

_____., 1993b, "Spatially Resolved Spectra of the Unidentified Infrared Features around HD 44179 (the Red Rectangle)," *ApJ*, **409**, 412.

Abstracts from conference proceedings published 1 April 1992 — 30 June 1993

Sloan, G.C. 1992, "Spatially Resolved 10 μm Spectra of Circumstellar Material around Evolved Stars," *BAAS*, **24**, 1302.

Sloan, G.C., Tandy, P.C., Pirger, B.E., & Hodge, T.M. 1993, "Spatial Structure in the Infrared Spectra of Three Evolved Stars," *BAAS*, **25**, 876.

LeVan, P.D., Sloan, G.C., & Little-Marenin, I.R. 1993, "Sources of the 13 μm Emission Feature Associated with Silicate Dust," *BAAS*, **25**, 877.

Spatially Resolved 10 μm Spectra of Circumstellar Material around Evolved Stars

G.C. Sloan (Phillips Laboratory)

Using a long-slit 10 μm spectrometer and new data analysis techniques, I have resolved and studied the circumstellar emission from two evolved stars, α Orionis and HD 44179. Maximum entropy reconstruction of α Ori reveals that emission from the object arises from two sources, an unresolved photosphere and an extended region of silicate dust emission. Reconstructions of HD 44179 resolve it into a cool central continuum source embedded within an extended region emitting the unidentified infrared (UIR) emission features at 7.7, 8.6, 11.3, and 12.7 μm. As the distance from the central source increases, the strength of the 7.7 μm feature decreases with respect to the 8.6 and 11.3 μm features, and there is an enhancement in the strength of the 11.3 μm emission roughly 1.5 arcsec from the central source. These results indicate that the UIR carriers are forming roughly 0.5 arcsec from the central source and evolving as they move outward.

Sources of the 13 μm Emission Feature Associated with Silicate Dust

P.D LeVan, G.C. Sloan (Phillips Lab.), I.R. Little-Marenin (Wellesley and Colorado)

We have carefully searched the Atlas of Low-resolution Spectra (IRAS Science Team 1986, *A&A Suppl.*, **65**, 607; Volk and Cohen 1989, *AJ*, **98**, 931) for sources of the 13 μm emission feature associated with silicate emission at 10 μm first discussed by Little-Marenin and Little (1988, *ApJ*, **303**, 305). We have identified 73 spectra for which the 13 μm feature is detected at 4σ or better. In several cases, we have confirmed the existence of the feature using GLADYS, the Air Force long-slit 10 μm spectrometer, at the Wyoming Infrared Observatory.

Our sample of 13 μm emission sources are nearly all late M giants, the majority of type M6 or M7. Variability types exist for 54; over half are SRb variables, while the remainder are fairly evenly divided among Miras and classes SRa and Lb. Most of the sources have LRS characterizations of 14, 15, or 21-24, i.e. weak silicate emission at 10 μm. The shape of the silicate emission feature varies from a nearly classic 10 μm profile broadened at longer wavelengths to a double-humped profile with peaks at both 10 and 11 μm. In the scheme of Little-Marenin and Little, these shapes would be characterized as Sil+, Sil++, and 3-component. The root mean square radial velocity of our sample is 31 km/sec. The mean angle from the galactic plane is 29 degrees, and there are no obvious dependencies with galactic longitude. These properties indicate that our sample consists of old Population I AGB stars.

We have also investigated how the strength of the 13 μm emission varies with spectral type, class and period of variability, LRS characterization, 10 μm feature width, and galactic position. We find no correlation with any of these properties, indicating that the 13 μm emission is not unique to any specific class of Population I AGB stars. There is no strong evidence for the contention that the 13 μm sources are precursors to S stars.

Spatial Structure in the Infrared Spectra of Three Evolved Stars   •

G.C. Sloan, P.C. Tandy (Phillips Lab.), B.E. Pirger (Cornell), T.M. Hodge (Wyoming)

We have spatially resolved three evolved sources using GLADYS, a long-slit 10 μm spectrometer, at the Wyoming Infrared Observatory. These observations, made in 1993 March, were the first for GLADYS after a complete replacement of the detector drive electronics, ADCs, and hardware co-adder. We studied each source in a north/south and an east/west slit orientation. For each set of observations, we fit a gaussian to the spatial profile at each wavelength to create a spatiogram, or plot of the width of the spectrum as a function of wavelength.

In both slit orientations, the spatiogram of α Orionis is widest at 10 μm, where the contribution from the silicate dust in the circumstellar shell is strongest. The FWHM at 10 μm is 2.0 arcsec, while our point-source comparison has a FWHM of 1.6 arcsec. These results are very similar to those presented for a N/S slit by Grasdalen, Sloan, and LeVan (1992, *ApJ*, 384, L25). IRC+10216 is also resolved in both slit orientations, having a FWHM of 1.9 arcsec at 11 μm, compared with 1.5 arcsec for a point source. No spectral structure is apparent in the spatiograms, indicating that there is little change in the spectral character of the emission across the source. AFGL 2688 (the Cygnus Egg) is clearly resolved in the N/S slit orientation, where its FWHM at 11 μm is 2.2 arcsec, but its spatiogram in the E/W slit orientation is barely distinguishable from that of a point source.

Abstract submitted for the 182st meeting of the AAS, Berkeley, CA

AIR FORCE GEOPHYSICS SCHOLAR PROGRAM

Sponsored by the

PHILLIPS LABORATORY

GEOPHYSICS DIRECTORATE

conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

GAS-PHASE ACIDITIES AND ELECTRON ATTACHMENT PROCESSES OF TRANSITION-METAL HYDRIDES

Prepared by: Dr. Amy E. Stevens-Miller

Research Location: Phillips Laboratory
                                     Geophysics Directorate
                                     Hanscom AFB, MA 01731

AF Research Colleague: Dr. John F. Paulson

Contract Number: F19628-86-C-0224

GAS-PHASE ACIDITIES AND ELECTRON ATTACHMENT PROCESSES OF TRANSITION-METAL HYDRIDES

by

Amy E. Stevens Miller

ABSTRACT

The gas-phase acidities of thirteen transition-metal hydride complexes were determined by bracketing or equilibrium proton-transfer reactions with reference anions and their conjugate acids. All the complexes examined are strong gas-phase acids, with several superacids, comparable in acidity to triflic acid. Electron attachment to the superacids shows the rates to be about 1/10 collisional, but increasing with temperature. The electron attachment rate coefficients to the carbonyl hydrides and $Ni(PF_3)_4$ are near the collisional limit. Electron attachment to the perfluoro compounds SF_6, SF_4, PF_3, PF_5, NF_3, and WF_6 was also studied. These complexes show a range of behavior in rate coefficients and in temperature dependence; neither rate coefficients nor the activation enthalpies show any obvious dependence on the electron affinity of the compound nor exothermicity of a dissociative attachment channel. Additional studies include determinations of the electron affinities of SF_4 and PF_5, and examination of the ion chemistry of Fe^-, C_2^- and HC_2^-, C_2O^-, and the $OH^- + NO$ system.

ACKNOWLEDGEMENTS

I.   INTRODUCTION:

A variety of Air Force applications require the control of electron density in plasmas, such as found in the atmosphere or in engine exhausts. The gas-phase acidity of a compound can be one indication of its potential as an electron scavenger. This is illustrated for a molecular hydride, "MH", whose gas-phase acidity (or more strictly, the enthalpy for the dissociation into the free anion and free proton) is given by eq (1),

$$\Delta H^{\circ}_{acid}[MH] = D[M-H] - EA[M] + 313.6 \text{ kcal/mol.} \qquad (1)$$

In eq (1), $D[M-H]$ is the M–H homolytic bond enthalpy, $EA[M]$ the electron affinity of the M radical, and the ionization energy of hydrogen, I.E.[H] is 313.6 kcal/mol. Weak gas-phase acids, e.g., CH_4 or NH_3, will be relatively difficult to dissociate into the anion and proton, and have large values for ΔH°_{acid}. Eq (1) shows this can be viewed in terms of a large difference between $D[M-H]$ and $EA[M]$. Strong gas-phase acids, e.g., HI or CF_3SO_3H, are much easier to dissociate into the anion and proton, and have small values for ΔH°_{acid}, corresponding to a small difference between $D[M-H]$ and $EA[M]$.

The gas-phase acidity of a compound is related to attachment of a thermalized electron, eq (2), for which the enthalpy of the attachment process is given in eq (3),

$$MH + e^- \text{ (thermal)} \longrightarrow M^- + H. \qquad (2)$$

$$\Delta H^{\circ}_{attachment} = D[M-H] - EA[M] - 1.48 \text{ kcal/mol.} \qquad (3)$$

As can be seen in eq (3), there is a slight enthalpy provided by the free electron at any temperature; in eq (3), the thermal enthalpy of the electron at 298 K, or 1.48 kcal/mol is used illustratively. As in eq (1), the important energetic quantity is the difference in $D[M-H]$ and $EA[M]$: if $D[M-H] \leq EA[M] + 1.48$ kcal/mol, the electron attachment process of eq (2) is exothermic. An examination of eqs (1) and (3) shows that this energetic requirement for exothermicity of eq (2) is equivalent to a requirement for the gas-phase acidity, $\Delta H^{\circ}_{acid} \leq 315.1$ kcal/mol.

The utility of these strong acids can be found in the ability to attach an electron to form a closed-shell, and hopefully unreactive, negative ion. The search for strong acids prompted our examination of a number of transition-metal hydride complexes, which we expected to be strong acids based on solution acidities[1] and known thermochemistry.[2]

II.   OBJECTIVES OF THE RESEARCH EFFORT:

Prior to my tenure as an Air Force Geophysics Scholar, I synthesized several transition-metal hydride complexes and made preliminary determinations of their gas-phase acidities at the Geophysics Laboratory.[3] We found the hydrides to be strong

acids or gas-phase superacids. The objective of my Geophysics Scholar research effort was to provide more accurate acidity determinations for those compounds, synthesize additional metal hydrides and determine their acidities, and determine the kinetics and products of the thermal electron attachment to the compounds. Although the hope was that the study of the electron attachment reactions themselves would be the top priority, this study required the moving of the Flowing Afterglow/Langmuir Probe (FALP) apparatus from Oklahoma, and turning into a "working" experiment by addition of a substantial amount of additional equipment, optimization of the apparatus (e.g., inlet design), and incorporation of new software for data analysis (installed in the Fall 1992 by Patrick Spanel and David Smith). Only then were we able to concentrate on how to prepare samples and obtain accurate attachment rate data, particularly for compounds with attachment rates near the theoretical limit.

A second thrust of the research was to have been an examination of the reactions of the metal anions produced by eq (2) with species important in atmospheric chemistry (e.g., O_2, NO, O_3, H_2O), and in combustion mixtures (e.g., O, H). No progress was made on this objective.

Other objectives of the work evolved during the period of the Geophysics Scholar tenure. One objective became a general understanding of both dissociative and non-dissociative electron attachment processes, particularly with respect to perfluorinated compounds (e.g., SF_4, PF_5). In related studies, I participated in electron affinity determinations for SF_4 and PF_5.[4, 5] We also examined attachment to non-superacids [$HMn(CO)_5$, $HRe(CO)_5$], complexes with only trifluorophosphine ligands [$Ni(PF_3)_4$], and the deuterated complex $DCo(PF_3)_4$. I also contributed to studies on other anions, including Fe^-,[6] C_2^- and HC_2^-,[7] C_2O^-,[8] and the $HO^- + NO$ system.[9]

III. GAS-PHASE ACIDITIES OF TRANSITION-METAL HYDRIDES

a. Methods. The preparation of the metal complexes was often the slow step in the research. In order to obtain facilities for the preparations, particularly the requirement for a good fume hood, John Paulson arranged for laboratory space for me in the Chemistry Department at Boston College. In addition to my vacuum/inert gas manifold and synthetic equipment, I had access to departmental instruments, all of which were crucial to some aspect of the syntheses. All metal hydride complexes were prepared by methods reported in the literature.[3] Goals of the research were roughly met, in that the compounds chosen for synthesis allowed study of successive PF_3 substituents, some middle to late transition-metals, and some examples of first versus second and third-row metal substitution, and the effect of other ligand substitutions [cyclopentadienyl (Cp), NO].

The rate constants and ion products for the ion-molecule reactions studied here were measured using a selected-ion flow drift tube (SIFDT), at the Geophysics Laboratory, which has been well-described elsewhere.[10] Some detail on handling these particular chemical systems follows.

For the $CpM(CO)_3H$ complexes the SIFDT instrumentation was not adequate, as it does not have the necessary resolution to study these high mass complexes for which the central metal has extensive isotopes. For this reason, we used an afterglow reactor, which is equipped with an Extrel quadrupole and associated electronics, currently operating to ~600 amu with unit mass resolution. The disadvantage of this system is that the ionizer is located in the upstream end of the flow tube, rather than in a remote source with mass selection of the primary ion. In general, we were able to select reagents which gave the reference anion as the major ion, and for which the metal anion produced by proton abstraction did not go on to react with the neutral used to create the reference anion. This instrument was designed for examining the kinetics of electron attachment reactions, and is equipped with a Langmuir probe. This probe proved crucial for the success of this work, as we found that if free electrons were allowed to persist in the flow system, erroneous kinetics and products were observed on the addition of the neutral metal reagent. In all cases the electron density was determined to be negligible at the neutral inlet prior to the experiments. Often the reagent used to create the reference anion could be chosen or adjusted to achieve complete electron loss, or on occasion O_2 was added prior to the microwave to alter the chemistry in such a way as to remove free electrons.

Generally the metal complexes were amenable to study. The percarbonyl complexes tend to be air, heat, and light sensitive, but the trifluorophosphine complexes, with even partial substitution, are generally quite stable. All complexes were protected from light while in use. The complexes do not exhibit the redox chemistry which makes most main-group acids, such as fluorosulfonic or trifluoroacetic, quite caustic. If wet, the PF_3 complexes do etch glass slowly--presumably by slow hydrolysis of the PF_3 ligands to make HF. No difficulties were found in handling the pure, dried complexes.

To obtain the kinetics of ion molecule reactions it is necessary to know the concentration of the neutral reagent, which is done by measuring the reactant mass flow. The reactant flow was measured by using a heat-transfer type flow meter (MKS), calibrated for N_2. A conversion factor for other gases depends on the heat capacity of the flowing gas. The conversion factors for the metal hydrides were estimated from the heat capacity for gaseous $Ni(CO)_4$, from which a conversion factor of 0.187 for $Ni(CO)_4$ relative to N_2 is derived. A second method to derive a conversion factor was to take a known mixture of $HMn(CO)_5$ or $HRe(CO)_5$ in argon, and determine the variation in the

"apparent" rate constant as a function of mixture percent. These tests gave a conversion factor of 0.20 ± 0.03, which was then used for all the metal hydrides. The inherent calibration errors suggest the rate constants reported here are generally accurate to only $\pm 35\%$, although rate constants are reproducible to with 5-10%. Rates reported for a series of ion reactants with a single metal hydride have a relative error of only $\sim 10\%$, since most of the overall error pertains to the flow calibration. Errors for rate constants reported for $HMn(CO)_4(PF_3)$ and $HMn(CO)(PF_3)_4$ have errors $\sim 50\%$, since they depend on additional factors, including the rate constants for $HMn(CO)_5$ and $HMn(PF_3)_5$, and any mass discrimination of the mass spectrometer used for determination of the component ratios in the mixtures used to study these complexes.

The $CpM(CO)_3H$ complexes have very low vapor pressures, which necessitated their addition as the vapor from powdered samples held in a glass bubbler with argon as a carrier gas. This method can be used to determine the concentration of the entrained complex, if the vapor pressure of the complex is known, and the pressure of the carrier gas determined (at each flow rate). We did not have a system capable of measuring the vapor pressures of these complexes at the time the data were taken, and will not report the relative rates of proton abstraction from the hydrides at this time. The 18-electron metal anions were readily produced by adding the entrained complexes just after the microwave ionizer, and rates of their reaction with reference acids determined and reported here.

All the reference acids introduced as reactant neutrals are quite caustic, and this necessitated frequent cleaning of the reactant flow meters. $(CF_3SO_2)_2NH$ has a very low vapor pressure (~ 0.1-0.2 Torr), and was introduced by passing a flow of argon over the solid sample held in a glass bubbler located after the argon flowmeter. HI was purified by trap-to-trap distillation prior to use. Flowmeter calibration factors for the neutral acids were determined by known or estimated heat capacities.

The halide ions were produced by dissociative electron attachment to methyl halides or trifluoromethyl halides; other anions were generated from the acid, anhydride, or ester; PO_3^- was generated from dimethyl phosphite; FSO_3^- from a 3:1 mixture of SF_4 (or SF_6) and SO_2. The $(CF_3SO_2)_2NH$ was a gift from D. DesMarteau; all other reagents were purchased from commercial sources and used as supplied.

$Mn_2(CO)_{10}$ and $ReMn(CO)_{10}$ were introduced into the source by passing argon or helium through the powdered sample held in a glass bubbler located after the flow meter; these complexes were used to produce $Mn(CO)_5^-$ and $Re(CO)_5^-$, respectively. The metal hydrides were introduced as a few percent mixture in helium or PF_3. All the metal complexes decompose in the source, although the dilution helped to minimize the speed at which the source stopped operating due to deposits of conducting metal.
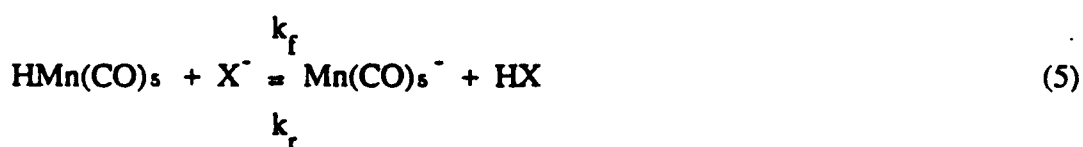
Table 1. Rate constants (in units of 10^{-9} $cm^3 \cdot molecule^{-1} \cdot s^{-1}$) for proton-transfer reactions of negative ions with transition-metal hydrides less acidic than HI.

reactant ion	reactant neutral		
	$HMn(CO)_5$	$HRe(CO)_5$	$HMn(CO)_4(PF_3)$
F^-	2.1	2.1[a]	-
CCl_3^-	1.1	-	-
HCO_2^-	-	1.1	-
NO_2^-	-	1.1	-
Cl^-	1.5	1.5	0.92
$CHF_2CO_2^-$	-	0.3	-
$CF_3C(O)CHC(O)CH_3^-$	0.16	NR (k<0.01)	-
$CHCl_2CO_2^-$	-	NR (k<0.01)	-
Br^-	1.0	NR (k<0.01)	0.74
$CF_3CO_2^-$	1.1	-	-
$CCl_3CO_2^-$	1.0	-	-
$CF_3C(O)S^-$	0.15	-	0.54
$CF_3C(O)CHC(O)CF_3^-$	NR (k<0.01)	-	0.15
I^-	NR (k<0.001)	NR (k<0.001)	0.12
PO_3^-	-	-	< 0.03
FSO_3^-	-	-	NR (k<0.05)

[a] Ion products are $Re(CO)_5^-$ (65%) and $HRe(CO)_3F^-$ (35%).

b. **Results.** Rate coefficients for reactions of negative ions with the transition-metal hydrides are reported in Tables 1 and 2.

Highly accurate relative acidities can be determined from the ratio of the forward and reverse rate coefficients, k_f/k_r, for a particular bimolecular proton transfer, as illustrated in eq (5) for reaction of $HMn(CO)_5$ with a reference base "X^-", giving the

$$HMn(CO)_5 + X^- \overset{k_f}{\underset{k_r}{\rightleftharpoons}} Mn(CO)_5^- + HX \qquad (5)$$

Table 2. Rate constants (in units of 10^{-9} $cm^3 \cdot molecule^{-1} \cdot s^{-1}$) for proton-transfer reactions of negative ions with transition-metal hydrides more acidic than HI.

reactant ion	reactant neutral		
	$HMn(CO)x(PF_3)_{6-x}$ [a]	$HMn(CO)(PF_3)_4$	$HMn(PF_3)_5$
Br^-	0.62	1.07	-
I^-	0.46	0.71	0.31
PO_3^-	0.62	0.98	0.47
FSO_3^-	0.49	0.93	0.31
$CF_3SO_3^-$	NR (k<0.002)	NR (k<0.005)	NR (k<0.005)
$(CF_3SO_2)_2N^-$	-	-	NR (k<0.01)

reactant ion	reactant neutral		
	$HCo(PF_3)_4$ [b]	$HRh(PF_3)_4$	$HIr(PF_3)_4$
Br^-	-	-	0.25
I^-	0.56	0.39	0.15
PO_3^-	0.69	0.53	0.4
FSO_3^-	0.62	0.44	0.19
$CF_3SO_3^-$	0.58	0.29	0.0003
$(CF_3SO_2)_2N^-$	NR (k<0.01)	NR (k<0.01)	NR (k<0.01)

[a] Sample consisted of a mixture of 78% $HMn(CO)_3(PF_3)_2$ and 22% $HMn(CO)_2(PF_3)_3$.

[b] Rate constant for reaction of Cl^- with $HCo(PF_3)_4$ is 1.0×10^{-9} $cm^3 \cdot molecule^{-1} \cdot s^{-1}$.

equilibrium constant, $Keq = k_f/k_r$, and hence the free energy change, $\delta\Delta G^{\circ}_{acid}$, eq (6),

$$\delta\Delta G^{\circ}_{acid} = -RT \ell n(k_f/k_r) = \Delta G^{\circ}_{acid}[HMn(CO)_5] - \Delta G^{\circ}_{acid}[HX]. \tag{6}$$

Rate constants for reactions of $Mn(CO)_5^-$ and $Re(CO)_5^-$ are reported in Tables 3 and 4, with the $\delta\Delta G^{\circ}_{acid}$ calculated using the forward rate constants from Table 1.

Table 3. Rate constants for reaction of $Mn(CO)_5^-$ with acids, and acidity of $HMn(CO)_5$ relative to the reference acid.

acid	rate constant $cm^3 \cdot molecule^{-1} \cdot s^{-1}$	$\delta\Delta G^{\circ}_{acid}$ kcal/mol
HBr	1.0×10^{-12}	-4.1
CF_3CO_2H	3.4×10^{-11}	-2.0
$CF_3C(O)SH$	8.9×10^{-10}	+1.1
$CF_3C(O)CHC(OH)CF_3$	1.5×10^{-10}	>+1.6
HI	2.7×10^{-10}	>+3.3
$HCo(PF_3)_4$	1.7×10^{-10}	--

Table 4. Rate constants for reaction of $Re(CO)_5^-$ with acids, and acidity of $HRe(CO)_5$ relative to the reference acid.

acid	rate constant $cm^3 \cdot molecule^{-1} \cdot s^{-1}$	$\delta\Delta G^{\circ}_{acid}$ kcal/mol
HCl	4.0×10^{-12}	-4.3
CHF_2CO_2H	3.7×10^{-10}	+0.2
HBr	8.0×10^{-10}	>+2.6
CF_3CO_2H	5.6×10^{-10}	>+2.4
$CF_3C(O)SH$	1.7×10^{-9}	--

Although the PF_3-substituted hydride complexes all give yields of the 18-electron anions on electron attachment in the source, the reference acids needed for examining the reverse reactions of the strong acids are generally not amenable to study. We chose to examine the reaction of $(CF_3SO_2)_2NH$ with only one of the metal anions:

$$Co(PF_3)_4^- + (CF_3SO_2)_2NH \rightarrow (CF_3SO_2)_2N^- + HCo(PF_3)_4 \quad (50\%) \quad (7a)$$

$$\rightarrow m/e \sim 559 + neutrals \ (50\%) \quad (7b)$$

Pathway (7b) is tentative, as the poor mass resolution did not allow unequivocal identification of the ion product. [We expect to check the product mass on the FALP.] The rate constant for reaction (7) is about a factor of 10^3 smaller than the rate constant for

reaction of Cl^- with $(CF_3SO_2)_2NH$, suggesting $k_7 \cong 10^{-12}$ $cm^3 \cdot molecule^{-1} \cdot s^{-1}$. Observation of the proton abstraction from $(CF_3SO_2)_2NH$, eq (7a), is consistent with the lack of observation of the forward reaction of $(CF_3SO_2)_2N^-$ with $HCo(PF_3)_4$, and thus confirms the bracketed position of the $HCo(PF_3)_4$ acidity (as given in Table 6, below).

In an attempt to find the relative acidities of the $HCo(PF_3)_4$ and $HRh(PF_3)_4$ complexes, reactions (8) and (9) were examined.

$$Co(PF_3)_4^- + HRh(PF_3)_4 \rightarrow NR \ (k < 0.01 \times 10^{-9} \ cm^3 \cdot molecule^{-1} \cdot s^{-1}) \qquad (8)$$

$$Rh(PF_3)_4^- + HCo(PF_3)_4 \rightarrow NR \ (k < 0.01 \times 10^{-9} \ cm^3 \cdot molecule^{-1} \cdot s^{-1}) \qquad (9)$$
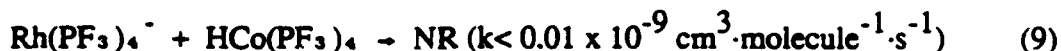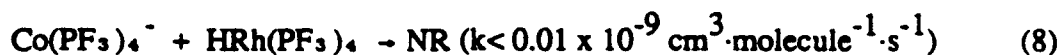
Other than the two reactions (8) and (9), and the reaction of $Mn(CO)_5^-$ with $HCo(PF_3)_4$ (Table 3), no other attempts were made to specifically examine proton-transfer reactions between any of the 18-electron metal anions and metal hydrides. Some indirect evidence for the fact that these reactions are generally slow comes from the examination of the $HMn(CO)_m(PF_3)_n$ mixtures. For example, reaction of reference anions Cl^- and Br^- with the $HMn(CO)_5/HMn(CO)_4(PF_3)$ mixture produces both $Mn(CO)_5^-$ and $Mn(CO)_4(PF_3)^-$ in the flow tube. If reaction (10) were to occur, as

$$Mn(CO)_5^- + HMn(CO)_4(PF_3) \rightarrow HMn(CO)_5 + Mn(CO)_4(PF_3)^- \qquad (10)$$

might be expected based on the ~ 4 kcal/mol exothermicity of proton transfer (Table 6, below), the intensity of $Mn(CO)_4(PF_3)^-$ would increase relative to the intensity of $Mn(CO)_5^-$ either with an increase in the flow of the $HMn(CO)_5 / HMn(CO)_4(PF_3)$ mixture or with an increase in reaction time (reaction distance in the flow tube). Changing product-ion intensity ratios with neutral flow or reaction distance were not observed with this mixture or any of the other $HMn(CO)_m(PF_3)_n$ mixtures which we examined, an indication that proton abstraction reactions such as eq (10) are at least an order of magnitude slower than the proton abstraction by the reference anions.

Rate constants for reactions of $CpCr(CO)_3^-$, $CpMo(CO)_3^-$, and $CpW(CO)_3^-$ with reference acids are given in Table 5. These experiments are consistent with results based on proton abstraction from the hydrides using the conjugate bases.

In the "strong acid" region, which could be considered from HCl and below, most of the determinations of quantitative acidities have been made very recently, by Kebarle and co-workers,[11, 12] Taft, Koppel, Anvia and co-workers,[13] Viggiano and Henchman and co-workers.[14, 15] Because the acidity scale is not as well established as for less acidic compounds (see Lias et al.[16]), we summarize the current best available set of ΔG°_{acid} in Table 6, with appropriate references. All ΔG°_{acid} of the reference compounds which were determined by equilibrium methods are assigned error limits of ± 2.0

Table 5. Rate constants (in units of 10^{-9} $cm^3 \cdot molecule^{-1} \cdot s^{-1}$) for reactions of $CpM(CO)_3^-$ with reference acids.

	reactant ion		
acid	$CpCr(CO)_3^-$	$CpMo(CO)_3^-$	$CpW(CO)_3^-$
HCl	NR (k < 0.005)	-	NR (k<0.005)
CHF_2CO_2H	-	-	0.15
HBr	NR (k < 0.005)	0.0054	0.52
CF_3CO_2H	0.099	0.087	0.20
$CF_3C(O)SH$	1.04	0.57	1.13
$CF_3C(O)CHC(OH)CF_3$	0.83	0.78	-

kcal/mol. Errors on the relative acidities, or $\delta\Delta G^\circ_{acid}$, are typically $\pm$ 0.2 kcal/mol, as indicated by comparison of data from several laboratories. Table 6 includes the enthalpy change, ΔH°_{acid}, for the proton loss process for each of the reference acids.

With the acidities for the reference acids established, we can now quantify the acidities of the transition-metal hydrides. $\Delta G^\circ_{acid}[HMn(CO)_5]$ is determined by comparison to HBr, CF_3CO_2H, and $CF_3C(O)SH$ (as shown in Table 3), and the average value reported in Table 6. For $HRe(CO)_5$, only $\delta\Delta G^\circ_{acid}$ by comparison to CHF_2CO_2H is known with any degree of accuracy (Table 4), and used to determine $\Delta G^\circ_{acid}[HRe(CO)_5]$ reported in Table 6.

In two cases we observe proton transfer rates in the forward direction which appear to depend on the nature of the reference anion: rate constants for reaction of $CF_3C(O)CH(O)CH_3^-$ with $HMn(CO)_5$, and reaction of $CF_3C(O)CHC(O)CF_3^-$ with $HMn(CO)_4(PF_3)$ are both about a factor of 10 below what is expected for collisional rate constants. These two reactions involve a highly delocalized "carbon" anion, and the slow rate of the reactions is consistent with the slow rate of reaction observed for other delocalized anions. In all other cases, reactions in the forward directions, proton abstraction from the metal hydrides by X^-, are all consistent with the usual assumptions used in bracketing reactions: Observation of fast proton transfer to the reference anion shows the metal hydride to be the stronger acid than the reference neutral, that is, $\Delta G^\circ_{acid}[\text{metal hydride}] \le \Delta G^\circ_{acid}[HX]$. Observation of a much slower reaction rate or no proton transfer shows the metal hydride is the weaker acid, that is, $\Delta G^\circ_{acid}[\text{metal hydride}]$

Table 6. Scale of acidities of the reference acids and transition-metal compounds. All data in kcal/mol.

acid	ΔH°_{acid}	ΔG°_{acid}	reference
HF	371.5 ± 0.2	365.7 ± 0.5	17
$CHCl_3$	357.1 ± 6.3	349.3 ± 6.0	16
HCO_2H	345.3 ± 2.3	338.3 ± 2.0	11
HNO_2	340.3 ± 2.1	332.6 ± 2.2	16
HCl	333.4 ± 0.2	328.0 ± 0.5	17
$HRe(CO)_5$	332.7 ± 3.6	324.0 ± 2.6	equilibrium
CHF_2CO_2H	330.8 ± 2.3	323.8 ± 2.0	11
$CF_3C(O)CHC(OH)CH_3$	328.8 ± 4.1	322.4 ± 2.0	12
$CHCl_2CO_2H$	328.9 ± 2.3	321.9 ± 2.0	11
HBr	323.5 ± 0.1	318.3 ± 0.4	17
CF_3CO_2H	324.4 ± 2.3	317.4 ± 2.0	11
$CpW(CO)_3H$	324 ± 5	317 ± 3	bracketed position
CCl_3CO_2H	319.9 ± 2.9	312.8 ± 2.0	16
$HMn(CO)_5$	323.5 ± 3.6	314.8 ± 2.6	equilibrium
$CpMo(CO)_3H$	321 ± 5	314 ± 3	bracketed position
$CpCr(CO)_3H$	321 ± 5	314 ± 3	bracketed position
$CF_3C(O)SH$	319.7 ± 3.0	313.6 ± 2.0	13
$CF_3C(O)CHC(OH)CF_3$	318.0 ± 4.1	311.6 ± 2.0	13
$HMn(CO)_4(PF_3)$	319.2 ± 5.0	310.5 ± 3.0	bracketed position
HI	314.4 ± 0.1	309.3 ± 0.4	17
HPO_3	310.7 ± 3.6	303.4 ± 4.2	14
FSO_3H	308 ± 3	301.3 ± 3.0	15
$HMn(CO)_3(PF_3)_2,$ $HMn(CO)_2(PF_3)_3,$ $HMn(CO)(PF_3)_4,$ $HMn(PF_3)_5$	309.4 ± 5	300.7 ± 3	bracketed positions
$HIr(PF_3)_4$	306.4 ± 5	300.7 ± 3	bracketed position
CF_3SO_3H	307 ± 3	300.1 ± 2.0	13
$HCo(PF_3)_4,$ $HRh(PF_3)_4$	304.0 ± 6	297.5 ± 4.6	bracketed positions
$(CF_3SO_2)_2NH$	302 ± 3	294.9 ± 2.0	13

$\geq \Delta G^{\circ}_{acid}[HX]$. For HCo(PF$_3$)$_4$ and the remaining hydrides, all the ΔG°_{acid} were determined by bracketing reactions in this fashion, with ΔG°_{acid} for the metal hydrides arbitrarily assumed to be the average of the acidities of the bracketing acids. For places where more than one metal complex is placed within a single bracketed position, as is particularly noticeable between FSO$_3$H and CF$_3$SO$_3$H, no relative ordering of the ΔG°_{acid} could be determined for the metal hydrides. No energetic ordering is therefore implied by the order in which the several metal complexes have been listed in the bracketed position.

Errors placed on the acidities of the metal hydrides incorporate the error in the absolute acidities and errors from the bracketing reactions or from $\delta\Delta G^{\circ}_{acid}$ determinations. Since the acidities of the metal hydrides are determined by bracketing experiments, however, the relative acidities of one to another are more accurate than the error bars indicate. All ΔH°_{acid} are determined by calculating the ΔS°_{acid} from rotational symmetry numbers and are given in Table 6.

c. Conclusions. The complexes HCo(PF$_3$)$_4$, HRh(PF$_3$)$_4$, HIr(PF$_3$)$_4$, HMn(CO)$_3$(PF$_3$)$_2$, HMn(CO)$_2$(PF$_3$)$_3$, HMn(CO)(PF$_3$)$_4$, and HMn(PF$_3$)$_5$ are "gas-phase superacids", comparable in acidity to triflic acid, and among the strongest acids known. All other hydride complexes examined are strong acids. The acidities increase from the middle to right of the transition series, and decrease with third-row metal (but not necessarily second-row metal) substitution. Data for the manganese series show that acidities increase ~4 kcal/mol with initial PF$_3$ for CO substitution, but the effect is evident only for the first two substitutions.

Some data on the bond strengths and electron affinities are available, although they will not be presented here.[3b, 2] These data indicate the transition-metal hydrides are generally strong gas-phase acids as a result of relatively small metal-hydrogen homolytic bond energies (~ 60 kcal/mol) and relatively large electron affinities (~ 2.5 eV), as given in eq (1). The acidity increase to the right in the periodic table is due to increasing electron affinities of the metal fragments, although some of the effect is due to the decreasing bond strengths. The decrease in acidity going down a column in the periodic table is due to an increase in the homolytic bond energies (with the electron affinities relatively constant). The acidity increases with PF$_3$ substitutions are due to increases in the electron affinities, with the bond energies approximately the same.

IV. ELECTRON ATTACHMENT TO TRANSITION-METAL COMPLEXES

a. Methods. Electron attachment reactions were studied using a flowing-afterglow Langmuir-probe (FALP) apparatus essentially identical to the apparatus constructed by

429

Smith and Adams,[18] and using the same data acquisition and analysis techniques. Several systematic errors may occur as a result of our chemical systems. Samples were introduced diluted to ~0.3% to 10% in helium, because the absolute sample flow rates needed were well below the limits of the flowmeters (0 to 10 sccm). An appropriate correction was applied to the flowmeter reading to account for the presence of the reactant gas. All the reactant vapors were found to be slightly "sticky"; we found it necessary to passivate the sample mixture bulbs with 1-10 torr of pure compound at least 8 hours prior to preparing any sample mixture. These difficulties from "stickiness"--absorption and/or adsorption onto the sample bulb and inlet walls are most pronounced with SF_4 (and to some extent PF_6). Passivation of the inlet system was accomplished by flowing pure SF_4 through the inlet overnight prior to data collection. The difficulties in handling these particular gases lead to large overall uncertainties in the attachment coefficients given in Tables 7 and 8; $\pm$ 35% would be a conservative estimate.

b. Results. $HCo(PF_3)_4$ was most extensively studied, since it is the strongest acid of the metal hydrides, and relatively easy to make in large (5 - 10 g) quantities. In order to study the role of the Co—H vibrational structure on the attachment, we also examined attachment to $DCo(PF_3)_4$ (made by H/D exchange with D_2O). Some data have been obtained on the electron attachment to $HRh(PF_3)_4$ [which is comparable in acidity to $HCo(PF_3)_4$, but a second-transition series metal], $HMn(CO)_5$ and $HRe(CO)_5$ [which are not superacids and lose only CO on attachment], and $Ni(PF_3)_4$ [which was chosen to learn more about the effect of the PF_3 ligand versus CO ligand]. A summary of the ion products and rate coefficients (β) of electron attachment are given in Table 7, and the rates are plotted as a function of temperature in Figure 1.

The activation enthalpies, $\Delta H^{\ddagger}$, for the attachment rates given for $HCo(PF_3)_4$ and $DCo(PF_3)_4$ are from fitting the entire temperature range to an exponential. The lowest temperature points correspond to $\Delta H^{\ddagger} \sim$ 60 meV, and the highest temperature points correspond to $\Delta H^{\ddagger} \sim$ 115 meV for the $HCo(PF_3)_4$ data shown. At the highest temperature, the ion $Co(PF_3)_4^-$ decreases to 22% of the ion intensity. There is some change for $HCo(PF_3)_4$ as compared to $DCo(PF_3)_4$ in the intensity of the ions produced by H/D loss vs. PF_3 loss. Finally, although the ΔH for the CO or PF_3 associative detachments are listed as unknown, they are believed to be nearly thermoneutral.

c. Conclusions. The proton transfer reactions of $HCo(PF_3)_4$ are "normal", that is, the proton is transferred at ~50% of the collisional rate to bases (Table 2). Thus, the fact that the electron attachment to $HCo(PF_3)_4$ proceeds with a rate coefficient approximately a factor of 10 below the maximum collisional rate constant is strong evidence that proton transfer is not the mechanism of dissociative electron attachment to $HCo(PF_3)_4$.[19]

Table 7. Summary of electron attachment processes for metal complexes.

compound	β (300 K) $cm^3 \cdot s^{-1}$	temperature dependence	ion products	thermochemistry
HCo(PF$_3$)$_4$	1.8×10^{-8}	$\Delta H^{\ddagger} \sim 88$ meV	Co(PF$_3$)$_4^-$ (41%)	$\Delta H = -11\ (\pm 6)$ kcal/mol
			HCo(PF$_3$)$_3^-$ (59%)	unknown
DCo(PF$_3$)$_4$	1.7×10^{-8}	$\Delta H^{\ddagger} \sim 85$ meV	Co(PF$_3$)$_4^-$ (31%)	$\Delta H = -9\ (\pm 8)$ kcal/mol
			DCo(PF$_3$)$_3^-$ (69%)	unknown
HRh(PF$_3$)$_4$	6×10^{-8}	decomposes with temperature	Rh(PF$_3$)$_4^-$ (55%)	$\Delta H = -11\ (\pm 6)$ kcal/mol
			HRh(PF$_3$)$_3^-$ (45%)	unknown
HMn(CO)$_5$	2×10^{-7}	not determined	HMn(CO)$_4^-$	unknown
HRe(CO)$_5$	2×10^{-7}	slight increase	HRe(CO)$_4^-$	unknown
Ni(PF$_3$)$_4$	1.7×10^{-7}	not determined	Ni(PF$_3$)$_3^-$	unknown
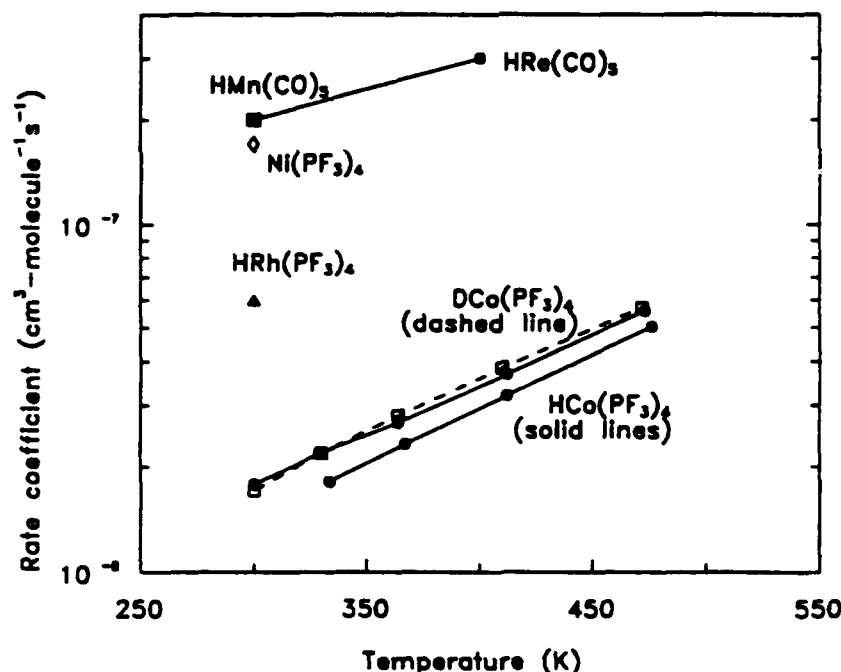
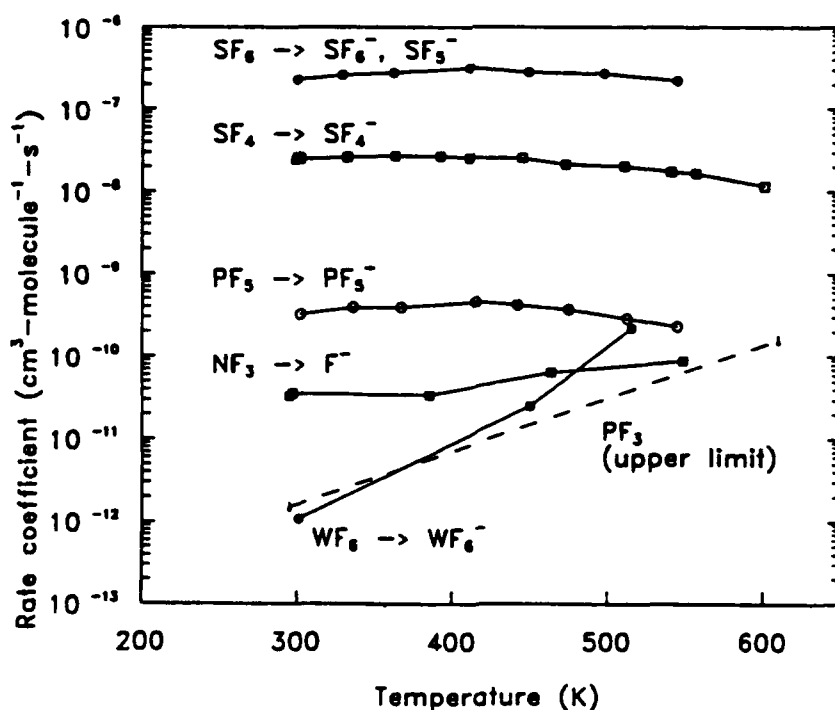FIGURE 1. Electron attachment rate coefficients for metal complexes.

HCo(PF$_3$)$_4$ and DCo(PF$_3$)$_4$ have identical attachment rate coefficients, although not ion fragmentation patterns, and the ion fragmentation patterns change with temperature. These facts suggest that the attachment proceeds through a long-lived resonant state, i.e., [HCo(PF$_3$)$_4^-$]*. HMn(CO)$_5$ and HRe(CO)$_5$ attach near the collisional rate; Ni(PF$_3$)$_4$ attaches slightly more slowly.

V.   ELECTRON ATTACHMENT TO PERFLUOROCOMPOUNDS

a. **Methods and Results.** I participated in a number of determinations of the electron attachment rates to perfluorocompounds. Methods used for the attachment studies are given in Section IV, and results are summarized in Table 8 and Figure 2.

Table 8. Summary of electron attachment processes to perfluorocompounds.

compound	β (300 K) $cm^3 \cdot s^{-1}$	temperature dependence	ion products	thermochemistry [16]
SF$_6$	2.3×10^{-7}	slight curvature	SF$_6^-$	$EA[SF_6] = 1.05\ (\pm 0.10)$ eV
			SF$_5^-$	$\Delta H[SF_5^-] \sim +5$ to $+28$ kcal/mol
SF$_4$	2.5×10^{-8}	slight curvature	SF$_4^-$	$EA[SF_4] = 1.5\ (\pm 0.2)$ eV [4]
PF$_5$	3.2×10^{-10}	slight curvature	PF$_5^-$	$EA[PF_5] \sim 0.80\ (\pm 0.20)$ eV [5]
PF$_3$	1.5×10^{-12}	NA	NA	$EA[PF_3] < 0$
NF$_3$	3.4×10^{-11}	$\Delta H^\ddagger \sim 50$ meV	F$^-$	$EA[NF_3] < 0$
				$\Delta H[F^-] = -23\ (\pm 3)$ kcal/mol
WF$_6$	1.1×10^{-12}	$\Delta H^\ddagger \sim 330$ meV	WF$_6^-$	$EA[WF_6] = 3.36\ (\pm 0.2)$ eV
				$\Delta H[WF_5^-] = +66\ (\pm 10)$ kcal/mol

Figure 2.   Electron attachment rate coefficients for perfluorocompounds.

b. Conclusions. Except for the obvious case of PF_3, which does not attach electrons due to lack of any exothermic pathway, there is no correlation in the attachment rates with electron affinity of the attacher or exothermicity of the dissociative attachment pathway. One supposition has been that SF_6 attaches rapidly due to the resonance with the near-thermoneutral associative detachment channel yielding SF_5^-. The ion thermochemistry is not available to test this hypothesis with the other purely associative attachers SF_4, PF_5, and WF_6. Electron scattering data as a function of electron energy are available for SF_6, SF_4, and PF_5, and may eventually provide insight into the negative ion states responsible for the attachment.

VI. OTHER STUDIES

I participated in a number of studies using the SIFDT instrumentation. These include: Determinations of the electron affinities of SF_4 (1.5 ± 0.2 eV) and PF_5 (~ 0.80 eV) by observation of charge transfer reactions.[4, 5] A study of the gas-phase reactions of Fe^- with five acids, which show that electron detachment is either a sole or dominant reaction channel; these reactions were used to determine $\Delta H_{acid}(FeH) = 345.2 \pm 4.4$ kcal/mol, which determines the bond energy $D^\circ_{298}(Fe-H) = 35.1 \pm 4.4$ kcal/mol.[6] A study of the chemistry of C_2^- and HC_2^- which shows both ions to be relatively unreactive, although the radical anion C_2^- reacts rapidly with radical neutrals.[7] A study of the reactivity of the radical anion C_2O^-, which shows it reacts by H, H^+, and H_2^+ abstraction, by nucleophilic displacement, charge transfer, and associative (or reactive) electron detachment; the chemistry is somewhat analogous to that of the radical anion O_2^-.[8] Finally, a study of HO^- shows it to react with NO by slow associative electron detachment ($k \sim 4 \times 10^{-12}$ $cm^3 \cdot molecule^{-1} \cdot s^{-1}$).[9] Results of these studies are all either published, in press, or in manuscript form, and will not be discussed in depth here.

VII. RECOMMENDATIONS

In general, the thermodynamic studies on acidities have been completed. Additional research should include the studies on the ions formed by electron attachment, including the study of the reactions with the parent neutrals and atmospheric and combustion species, as mentioned in Section II. Substantial work on the electron attachment reactions to the metal hydrides and related complexes remains.

In terms of implementation, the next step would be to form a collaboration to study the effect of these complexes on the plasma chemistry and physics on their addition to a plasma arc in a laboratory setting. Only then would I recommend studies involving atmospheric releases or lab or in situ engine exhaust modification.

REFERENCES

1. Kristjánsdóttir, S. S., J. R. Norton, "Acidity of Hydrido Transition Metal Complexes in Solution" In: Transition Metal Hydrides: Recent Advances in Theory and Experiment; Dedieu, A., Ed.; VCH: New York, 1990.

2. Simoes, J. A. M. and J. L. Beauchamp, "Transition metal–hydrogen and metal–carbon bond strengths: The keys to catalysis," Chem. Rev., Vol. 90, pp. 629-688, 1990.

3. a) Miller, A. E. S., A. R. Kawamura and T. M. Miller, "Effects of Metal and Ligand Substitutions on Gas-Phase Acidities of Transition-Metal Hydrides," J. Am. Chem. Soc., Vol. 112, pp. 457-458, 1990. b) Miller, A. E. S., T. M. Miller, "Gas-phase acidities of transition-metal carbonyl and trifluorophosphine hydrides," pp. 253-267 In: Energetics of Organometallic Compounds, Martinho Simoes, J. A., Ed.; Kluwer, Dordrecht, The Netherlands, 1992. c) Miller, A. E. S., T. M. Miller, G. K. Rekha, and A. R. Lambrecht, manuscript in preparation.

4. Viggiano, A. A., T. M. Miller, A. E. S. Miller, R. A. Morris, J. M. Van Doren and J. F. Paulson, "SF$_4$: electron affinity determinations by charge-transfer reactions," Int. J. Mass Spectrom. Ion Processes, Vol. 109 pp. 327-338, 1991.

5. Miller, T. M., A. E. S. Miller, A. A. Viggiano, R. A. Morris and J. F. Paulson, manuscript in preparation.

6. Miller, A. E. S., T. M. Miller, R. A. Morris, A. A. Viggiano, J. M. Van Doren and J. F. Paulson, "Reactions of Fe$^-$ with acids: gas-phase acidity and bond energy of FeH," Int. J. Mass Spectrom. Ion Processes, Vol. 123, pp. 205-216, 1993.

7. Van Doren, J. M., R. A. Morris, A. A. Viggiano, A. E. S. Miller, T. M. Miller and J. F. Paulson, "Chemistry of C$_2^-$ and HC$_2^-$ with nitrogen, oxygen and nitrogen oxides," Int. J. Mass Spectrom. Ion Processes, Vol. 117, pp. 395-414, 1992.

8. Van Doren, J. M., T. M. Miller, A. E. Stevens Miller, A. A. Viggiano, R. A. Morris and J. F. Paulson, "Reactivity of the Radical Anion OCC$^-$," J. Am. Chem. Soc., in press.

9. Van Doren, J. M., A. A. Viggiano, R. A. Morris, A. E. S. Miller, T. M. Miller, J. F. Paulson, C. A. Deakyne, H. H. Michels and J. A. Montgomery, Jr., "Experimental and theoretical study of the reaction of HO$^-$ with NO", J. Chem. Phys., in press.

10. Viggiano, A. A., R. A. Morris, F. Dale, J. F. Paulson, K. Giles, D. Smith and T. Su, "Kinetic energy, temperature, and derived rotaional temperature dependences for the reactions of Kr$^+$ (^{2}P$_{3/2}$) and Ar$^+$ with HCl," J. Chem. Phys., Vol. 93, pp. 1149-1157, 1990.

11. Caldwell, G., R. Renneboog and P. Kebarle, "Gas phase acidities of aliphatic carboxylic acids, based on measurements of proton transfer equilibria," Can. J. Chem., Vol. 67, pp. 611-618, 1989.

12. Cumming, J. B. and P. Kebarle, "Summary of gas phase measurements involving acids AH. entropy changes in proton transfer reactions involving negative ions. Bond dissociation energies D(A–H) and electron affinities EA(A)," Can. J. Chem., Vol. 56, pp. 1-9, 1978. Value adjusted to values for reference acids given in Caldwell et al.[11]

13. Taft, R. W., F. Anvia, I. A. Koppel, D. D. DesMarteau, L. M. Yagupolskii, Y. L. Yagupolskii, N. V. Kondratenko, A. Y. Volkonskii, V. M. Valsov, N. V. Ignat'ev, personal communication (1992). All values adjusted to $\Delta G^*_{acid}[CF_3 CO_2 H]$ = 317.4 kcal/mol from Caldwell et al.[11].

14. Viggiano, A. A., R. A. Morris, F. Dale, J. F. Paulson, M. J. Henchman, T. M. Miller and A. E. S. Miller, "Gas-phase acidities of HPO_3 and HPO_2: Enthalpies of deprotonation," J. Phys. Chem., Vol. 95, pp. 1275-1276, 1991.

15. Viggiano, A. A., M. J. Henchman, F. Dale, C. A. Deakyne and J. F. Paulson, "Gas-phase reactions of weak BrØnsted bases I^-, PO_3^-, HSO_4^-, FSO_3^-, and $CF_3 SO_3^-$ with strong BrØnsted acids $H_2 SO_4$, $FSO_3 H$, and $CF_3 SO_3 H$. A quantitative intrinsic superacidity scale for the sulfonic acids $XSO_3 H$ (X - HO, F, and CF_3)," J. Am. Chem. Soc., Vol. 114, pp. 4299-4306, 1992. Value relative to $CF_3 SO_3 H$ is adjusted to $\Delta G^*_{acid}[CF_3 SO_3 H]$ = 300.7 kcal/mol (see Table 6.)

16. Lias, S. G., J. E. Bartmess, J. F. Liebman, J. L. Holmes, R. D. Levin, W. G. Mallard, "Gas-Phase Ion and Neutral Thermochemistry", J. Phys. Chem. Ref. Data, Vol. 17, Suppl. No. 1., 1988.

17. Chase, M. W., Jr., C. A. Davies, J. R. Downey, Jr., D. J. Frurip, R. A. McDonald, A. N. Syverud, "JANAF Thermochemical Tables," J. Phys. Chem. Ref. Data, Vol. 14, Suppl. No. 1., 1985.

18. Smith, D., N. G. Adams, E. Alge, "Attachment coefficients for the reactions of electrons with CCl_4, $CHCl_3$, Cl_2 amd SF_6 determined between 200 and 600 K using the FALP technique," J. Phys. B, Vol. 17, 461-472, 1984.

19. Adams, N. G., D. Smith, A. A. Viggiano, J. F. Paulson and M. J. Henchman, "Dissociative attachment reactions of electrons with strong acid molecules," J. Chem. Phys., Vol. 84, pp. 6728-6731, 1986.